University of Alberta

Development of Quantitative Measures of Lumbar Disc Contour Abnormalities for MRI: Relialibility and Validity

by

Éric Parent ©

A thesis submitted to the final exam committee in partial fulfilment of
the
requirements for the degree of Doctor of Philosophy

in

Rehabilitation Science

Faculty of Rehabilitation Medicine

Edmonton, Alberta

Fall 2005

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-494-08710-2
Our file Notre référence
ISBN: 0-494-08710-2

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

ABSTRACT

Qualitative ratings of disc contour abnormalities (bulging and herniation) are sub-optimal for research on the determinants and effects of MRI findings. More precise measures would enhance studies of small effects. Qualitative ratings' limitations include lack of agreement on rating definitions, sub-optimal reliability, and rarity of validity evidence for ratings of contour abnormalities' magnitude. Quantitative measures of other imaging findings (signal intensity, stenosis) have helped address such limitations and are needed for research on contour abnormalities. This study's objectives were to develop and validate quantitative measures of disc contour abnormality on MRIs and to assess their measurement properties.

Area and distance measures of disc contour were developed. The measures were derived from a manually segmented outline of the vertebrae and discs, using two types of reference points: real and "theoretical corners" (as if there were no vertebra remodeling). Reliability studies were conducted to assess the influence of differences between repeated measurements, raters, image resolution, scanners, and spine positions. Convergent and divergent validity studies were conducted to examine the correlations amongst the different quantitative measures and with qualitative ratings.

The intra-evaluator, inter-evaluator, and inter-scanner reliability of the quantitative measures was adequate for group-level inferences and superior to that of qualitative ratings. The effect of image resolution on reliability was smaller (2.0% of error variance) than that due to intra-evaluator (3.0%), inter-evaluator (5.0%), and inter-scanner differences (8.3%). Posterior contour abnormalities were significantly smaller in

flexion and in extension than in the neutral position (8.4%-30.8%). For optimal reliability, the same MR scanner, imaging parameters, subjects' positioning, and evaluator should be used. Standardized positioning is particularly important for subjects suspected to have important contour abnormalities, because the differences between positions are larger in subjects with severe contour abnormalities. The relative magnitude of the correlations between quantitative measures and measures of the same construct or of different constructs supported this study's hypotheses and provided evidence of convergent and divergent validity.

At this stage, the quantitative measures provide a more reliable and precise measure of disc contour abnormalities than qualitative ratings and are recommended for research on the determinants and effects of variations in disc contour.

ACKNOWLEDGEMENTS

Cindy, I am fortunate that our paths crossed during this PhD and look forward to our future together. Your support, patience, and motivating power were most appreciated.

I want to thank my supervisors, Dr. Michele Crites Battié and Dr. Tapio Videman, who provided extensive feedback and support during this PhD project. I am grateful that they provided me with access to the rich Twin Spine Study database. They also encouraged me to pursue a number of enriching training opportunities such as teaching in different programs at the University of Alberta, training within the Alberta Provincial CIHR Training Program in the Bone and Joint Training Program, and participating in international conferences both as a presenter and a member of the planning committees.

I was also fortunate to have the support of a truly interdisciplinary supervisory committee who provided constructive feedback. I appreciated Dr. Todd Rogers' enthusiasm for the project and the time spent discussing statistical challenges. Dr. Richard Snyder's support was greatly appreciated during the planning stages of the resolution and flexion-extension projects. I am grateful to Dr. Richard Hu's flexibility in participating in the first reliability study and for providing valuable clinical insight. My committee members modeled effective interdisciplinary collaboration and inspire me to continue to pursue collaborative work in the future.

I also want to express my gratitude to Micheal Kennedy and Jennifer Klein, colleagues from the PhD program, for their tremendous personal support and challenging scientific discussions. Your friendship helped make my PhD experience enjoyable and a period of great personal and scientific growth. I'd like to thank my office mates and fellow trainees Doug, Rikka, Smaranda, Christie, Lino, Edgar, Tyler, and Troy; they all played a part in making my PhD experience both scientifically and culturally enriching.

I am also grateful to the following agencies for the excellent financial support I received during my PhD: fellowship from the Canadian Institute for Health Research, bourse de recherche en Santé du Fonds de la Recherche en Santé du Québec, Incentive award from th Alberta Heritage Foundation for Medical Research, University of Alberta fellowship and travel grant from the Alberta Provincial CIHR Training Program in Bone and Joint Health.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION, REVIEW OF THE LITERATURE AND OBJECTIVES 1

IMPROVING MEASURES OF DISC CONTOUR ABNORMALITY – A SHORT RATIONALE 1

EPIDEMIOLOGY OF LOW BACK PAIN 2

ETIOLOGY HYPOTHESES 3

The Intervertebral Disc as a Primary Suspect for LBP *4*

The Intervertebral Disc as a Source of LBP: Mechanisms 4

The Alternative Suspects - Lumbar Structures Causing LBP? *6*

IMAGING OF DEGENERATIVE CHANGES IN THE LUMBAR SPINE 8

Brief Historical Overview of Disc Imaging *8*

Magnetic Resonance Imaging *9*

Qualitative Findings on Lumbar Spine MR images **9**

Are Degeneration Findings Associated with LBP Symptoms? 12

Assessing Disc Degeneration on MR images – Not Only to Search for a Cause of LBP 15

Suboptimal Measurement Properties of Clinical Ratings 16

Quantitative Measures **20**

Expected Advantages of Quantitative Measures over Subjective Ratings 20

Quantitative Measures: What Progress Has Been Made so Far? 22

THESIS OBJECTIVES 27

REFERENCES - CHAPTER 1 32

CHAPTER 2: QUANTITATIVE MRI MEASURES OF LUMBAR DISC CONTOUR ABNORMALITY: DEVELOPMENT AND RELIABILITY 49

INTRODUCTION 49

MATERIALS AND METHODS 51

Subjects *51*

MR Imaging *51*

Quantitative Assessments *52*

Qualitative Assessments *53*

Analyses *54*

RESULTS 55
Differences in Reliability between Quantitative Measures *55*
Sources of Errors *55*
Subjective Ratings' Reliability *56*
DISCUSSION 56
Difference in Reliability *56*
Sources of Errors *57*
Other Sources of Error *58*
Reliability of Average Measures *58*
Comparing Qualitative and Quantitative Measures *59*
REFERENCES – CHAPTER 2 65

CHAPTER 3: MEASUREMENT ERROR DUE TO INTRA-EVALUATOR AND INTER-SCANNER DIFFERENCES IN QUANTITATIVE MEASURES OF LUMBAR DISC CONTOUR ABNORMALITY ON MR IMAGES 69

INTRODUCTION 69
MATERIALS AND METHODS 70
Subjects *70*
Imaging *71*
Quantitative Assessment *71*
Analyses *73*
RESULTS 73
Reliability Coefficients *74*
Difference in Reliability between Disc Levels *74*
Difference in Reliability between Measures at a Given Disc Level *74*
Measurement Errors *75*
Partitioning Error Variance *75*
DISCUSSION 75
Comparisons with the Existing Literature *77*
Implications and Recommendations *78*
REFERENCES – CHAPTER 3 84

CHAPTER 4: THE EFFECT OF LUMBAR FLEXION AND EXTENSION ON DISC CONTOUR ABNORMALITY MEASURED QUANTITATIVELY ON MRI 88

INTRODUCTION 88

MATERIAL AND METHODS 89

Subjects *89*

Imaging *90*

Quantitative Measures of Disc Contour Abnormality *90*

Analyses *91*

RESULTS 92

Angles *92*

Dimensions of Posterior Disc Contour Abnormality *92*

Dimensions of Anterior Disc Contour Abnormality *92*

Correlation between Difference in Quantitative Contour Abnormality between Positions and Contour Abnormality Severity in the Neutral Position *92*

Posterior Disc Contour Abnormality 92

Anterior Disc Contour Abnormality 93

DISCUSSION 93

The Effect of Flexion and Extension on Disc Contour Abnormality *94*

Degree of Degenerative Findings and Differences between Positions *95*

Study Strengths and Limitations *96*

REFERENCES – CHAPTER 4 105

CHAPTER 5: EVIDENCE OF CONVERGENT AND DIVERGENT VALIDITY FOR QUANTITATIVE MEASURES OF LUMBAR DISC CONTOUR ABNORMALITY ON MR IMAGES 109

INTRODUCTION 109

OBJECTIVES 111

HYPOTHESES 112

METHODS 113

Subjects *113*

MR Imaging *113*

Quantitative Assessment of MR Images *114*

Segmentation Procedure 115

Quantitative Measures 115

Qualitative Assessment of MR images *116*

Raters and Qualitative Assessment Definitions Used for Samples 1 and 2 . 116

Qualitative Data Entry 117

Qualitative Ratings' Reliability 117

Analyses *117*

Descriptive Analyses 117

Convergent and Divergent Validity Correlations 117

Convergent and Divergent Validity Correlations Using Midsagittal Qualitative Ratings 118

Convergent and Divergent Validity Correlations Using Midsagittal and Para-Sagittal Qualitative Ratings 119

Convergent and Divergent Validity Correlations Amongst the Quantitative Measures of Disc Contour Abnormality 120

Differences in Quantitative Measures of Contour Abnormality between Sagittal Slices 120

RESULTS 120

Subjects' Description, Frequencies of Qualitative Ratings and Mean Quantitative Measures *120*

Convergent-Divergent Validity Using Midsagittal Qualitative Ratings *121*

Orthopedic Surgeon, Using Ratings from Table 5.2, Sample 1 (Table 5.6). 121

Cross-Validation Using Data from Sample 1b *121*

Orthopedic Surgeon, Using Ratings from Table 5.2 (Table 5.7) 121

Orthopedic Surgeon, Using Definitions in Table 5.3: (Table 5.8) 121

First Radiologist, Using Ratings from Table 5.3 (Table 5.9) 122

Second Radiologist, Using Ratings from Table 5.4 (Table 5.10) 122

Convergent-Divergent Validity Using Midsagittal and Para-sagittal Qualitative Ratings *123*

Convergent-Divergent Validity Amongst Quantitative Measures *123*

Differences in Quantitative Contour Abnormality Dimensions Between Sagittal Slices *124*

DISCUSSION 125

Convergent and Divergent Validity with Midsagittal Qualitative Ratings *125*

Convergent-Divergent Validity with Midsagittal and Para-sagittal Qualitative Ratings *127*

Convergent and Divergent Validity Amongst Quantitative Measures *128*
Literature on the Validity of Quantitative Measures of Contour Abnormality *130*
Differences in Quantitative Contour Abnormality between Sagittal Slices *130*
Recommendations for Future Research *131*
CONCLUSIONS 133
REFERENCES – CHAPTER 5 153

CHAPTER 6: GENERAL DISCUSSION 158

SUMMARY OF THE MAIN FINDINGS 158
Reliability *158*
Effect of Position *159*
Convergent and Divergent Validity *159*
Differences in Contour Abnormalities Measurements between Midsagittal and Parasagittal Images *160*
STUDY LIMITATIONS 160
Distinguishing Bulging, Herniation, Protrusion and Extrusion Contained or not. *160*
Manual Segmentation is a Source of Error and Time-consuming *162*
STUDY STRENGTHS 163
FURTHER DEVELOPMENT WORK SUGGESTED FOR QUANTITATIVE MEASURES OF CONTOUR ABNORMALITY 165
Exploring Additional Sources of Unexplained Measurement Error *165*
Image Segmentation Procedure *166*
Providing Stronger Validity Evidence *167*
Which Slice(s) should Be Used to Measure Contour Abnormality Quantitatively? *167*
Study Alternative Measures of Disc Contour Abnormality *168*
POTENTIAL APPLICATIONS FOR QUANTITATIVE MEASURES OF DISC CONTOUR ABNORMALITY 169
Longitudinal Studies *169*
Studies on the Determinants and Epidemiology of Disc Contour Abnormality *170*
Studies of the Effect of Postures and Loading on Disc Contour Abnormality *171*
Studies of the Signs and Symptom, if any, Associated with Disc Contour Abnormality *172*
CONCLUSION 173
REFERENCES - CHAPTER 6 174

LIST OF TABLES

Table 1.1. Prevalence of disc-related degeneration findings on MR images of the lumbar spine in asymptomatic subjects 29

Table 1.2. Prevalence of disc-related degeneration findings on MR images of the lumbar spine in subjects with symptoms from case-control, population-based, or occupational group studies 30

Table 1.3. Prevalence of disc-related degeneration findings on MR images of the lumbar spine in population-based or occupational group samples selected without regards to low back pain 31

Table 2.1. Mean values, standard deviation (SD), and error of measurements for quantitative measures of disc contour abnormality for each level of the lumbar spine (n=31) 60

Table 2.2. Intra-, inter-evaluator generalizability coefficients, and 95% confidence intervals for the reliability for the quantitative measures of disc contour abnormality for each level of the lumbar spine (n=31) 61

Table 2.3. Mean values, standard deviation (SD), and delta error of measurements for qualitative measures of disc bulging/herniation for each level of the lumbar spine 62

Table 2.4. Generalizability coefficients estimates and 95% confidence intervals of the reliability for the qualitative measures of disc bulging/herniation on high-resolution images by any rater on any occasion for each level of the lumbar spine 63

Table 3.1 Mean values, standard deviation (SD), and error of measurement for quantitative measures of disc contour abnormality by any evaluator on any scanner for each level of the lumbar spine 80

Table 3.2. Generalizability coefficients (Gcoef) estimates and 95% confidence intervals of the reliability for the quantitative measures of disc contour abnormality by any evaluator on any MR scanner for each level of the lumbar spine 81

Table 3.3. Percent error variance due to inter-scanner and intra-evaluator differences at each lumbar disc level 82

Table 4.1. Mean disc segmental angle in the flexed, neutral, and extended positions 98

Table 5.1. Subjects' description 135

Table 5.2. Qualitative rating scales used by the orthopedic surgeon in Sample 1, 1b, and 2 136

Table 5.3. Qualitative rating scales used by the orthopedic surgeon and the first radiologist in Sample 1b and 2 137

Table 5.4. Qualitative rating scales used by the second radiologist in Sample 1b .. 138

Table 5.5. Reliability (ICC) of qualitative ratings in Sample 1b using definitions in Table 5.3 139

Table 5.6. Mean correlation over all lumbar disc levels between quantitative measures of disc contour abnormality and qualitative ratings using definitions from Table 5.2 by an orthopedic surgeon in a sample of 132 subjects (Sample 1) 140

Table 5.7. Mean correlation over all lumbar disc levels between quantitative measures of disc contour abnormality and qualitative ratings using definitions from Table 5.2 by an orthopedic surgeon in a sample of 88 subjects (Sample 1b) 141

Table 5.8. Mean correlation over all lumbar disc levels between quantitative measures of disc contour abnormality and qualitative ratings using definitions from Table 5.3 by an orthopedic surgeon in a sample of 88 subjects (Sample 1b) 142

Table 5.9. Mean correlation over all lumbar disc levels between quantitative measures of disc contour abnormality and qualitative ratings using definitions from Table 5.3 by the first radiologist in a sample of 88 subjects (Sample 1b) 143

Table 5.10. Mean correlation over all lumbar disc levels between quantitative measures of disc contour abnormality and qualitative ratings using definitions from Table 5.4 by the second radiologist in a sample of 88 subjects (Sample 1b) 144

Table 5.11. Mean correlation over all lumbar disc levels between the maximum quantitative measures of disc contour abnormality of the three most central sagittal images and qualitative ratings for the midsagittal images using definitions from Table 5.2 by the orthopedic surgeon in Sample 2 of 47 subjects 145

Table 5.12. Mean correlation over all lumbar disc levels between the maximum quantitative measures of disc contour abnormality of the three sagittal images corresponding to the right side of the lumbar spine and qualitative ratings for the right side images using definitions from Table 5.2 by the orthopedic surgeon in Sample 2 of 47 subjects 146

Table 5.13. Mean correlation over all lumbar disc levels between the maximum quantitative measures of disc contour abnormality of the three sagittal images corresponding to the left side of the lumbar spine and qualitative ratings for the left side images using definitions from Table 5.2 by the orthopedic surgeon in Sample 2 of 47 subjects 147

Table 5.14. Mean Pearson correlation coefficients of five lumbar disc levels between quantitative measures of disc contour abnormality from the midsagittal image for subjects in Sample 3 with normal disc contour (no moderate or severe bulge and no herniation) (n=212 to 259 depending on disc level) 148

Table 5.15. Mean Pearson correlation coefficients of five lumbar disc levels between quantitative measures of disc contour abnormality from the midsagittal image for subjects in Sample 3 with abnormal posterior disc contour (moderate or severe bulge or herniation posteriorly) (n=20 to 71 depending on disc level). 149

Table 5.16. Mean Pearson correlation coefficients of five lumbar disc levels between quantitative measures of disc contour abnormality from the midsagittal image for subjects in Sample 3 with abnormal anterior disc contour (moderate or severe bulge or herniation anteriorly) (n=24 to 63 depending on disc level)... 150

LIST OF FIGURES

Figure 2.1. Measures of disc contour abnormality, using real and theoretical corners 64

Figure 3.1. Measures of disc contour abnormality, using real and theoretical corners 83

Figure 4.1. Imaging position in neutral, flexion, and extension 101

Figure 4.2. Measures of disc contour abnormality, using real and theoretical corners 102

Figure 4.3. Example of a 57-year-old male with a large posterior contour abnormality in the neutral position and large decrease in quantitative disc contour abnormality in flexion or extension compared to neutral 103

Figure 4.4. Example of a 35-year-old male with a small posterior contour abnormality in the neutral position and small decrease in quantitative disc contour abnormality in flexion or extension compared to the neutral position 104

Figure 5.1. Measures of disc contour abnormality, using real and theoretical corners 151

Figure 5.2. Profile of contour abnormality dimensions for each level depending on which sagittal image is used to obtain measurement. Subject from Sample 2 qualitatively identified as having normal disc contour (left) compared to subjects known to have some levels with moderate or severe bulging (right). 152

LIST OF APPENDICES

APPENDIX A: The intervertebral disc – contextual information 181

APPENDIX B: Definitions of disc-related imaging findings from the Nomenclature and classification of lumbar disc pathology 183

APPENDIX C: An illustrated description of the manual segmentation procedure 186

APPENDIX D: Variance components for Chapter 2 191

APPENDIX E: Variance components for Chapter 3 198

APPENDIX F: Rationale for the hypotheses of convergent and divergent validity 201

APPENDIX G: Frequency distributions of the qualitative scores attributed by the different raters using the different rating definitions for the different samples. 205

APPENDIX H: Means of the different quantitative measures of disc contour abnormality for each sample 208

CHAPTER 1: INTRODUCTION, REVIEW OF THE LITERATURE AND OBJECTIVES

IMPROVING MEASURES OF DISC CONTOUR ABNORMALITY – A SHORT RATIONALE

Bulging and herniation, referred to hereafter as "disc contour abnormalities," have been the focus of much attention since Mixter and Barr's suggestion that deformation of a disc creating pressure on a nerve could produce low back pain (LBP).[1] Currently, the intervertebral disc is the focus of numerous diagnostic and treatment procedures, many of which are aimed at preventing, detecting, or correcting disc contour abnormalities.

However, several controversies surround the assessment of disc contour abnormalities with magnetic resonance imaging (MRI)

(1) A wide variety of rating definitions are used.

(2) Reliability estimates cover a wide range and are often sub-optimal for use in correlation studies or the study of small effects.

(3) Reliability and definitional problems may contribute to the controversy surrounding whether disc contour abnormalities are more prevalent in persons with symptoms than in asymptomatic subjects.

(4) Validity evidence for ratings of disc contour abnormalities dimensions is lacking.

(5) The commonly used ratings of disc contour appear to lack adequate sensitivity to change for detecting difference in longitudinal studies.

Developing a quantitative measure of disc contour abnormality may help clarify the aforementioned issues. Quantitative measures have already been developed for disc signal intensity and spinal stenosis with high reliability and the ability to detect small differences between repeated measurements.

The literature supporting this brief argument will be reviewed in this chapter. A summary of the epidemiology of LBP is presented. This summary shows that in most cases of LBP the etiology is not definitively known. Hypotheses on the etiology of non-specific LBP are discussed with emphasis on the intervertebral disc and its contour abnormalities as prime suspect. The mechanisms by which the intervertebral disc could be a source of pain are reviewed to demonstrate that in most cases the extent of disc contour abnormalities may be a contributing factor to LBP. Different imaging findings related to the intervertebral disc are defined, and studies on the prevalence of these findings in subjects with and without symptoms are reviewed to illustrate the controversy surrounding whether imaging findings are associated with pain or not. The qualitative ratings of disc contour abnormalities are reviewed, and the variability in the definitions and the limitations in these ratings' reliability, validity and ability to detect differences are discussed.

To address the limitations of qualitative measures, quantitative measures of disc contour abnormalities are presented as an alternative along with suggestions of how quantitative measures may assist research on disc contour abnormalities. The literature on quantitative measures is reviewed in order to shed light on the controversies outlined above. Finally, the introduction closes with the thesis objectives outlining the steps followed to develop and assess new quantitative measures of disc contour abnormalities for MRI.

EPIDEMIOLOGY OF LOW BACK PAIN

LBP is a costlyproblem[2-8] affecting 50% to 85%[8-16] of the population at some point in the lifetime. In a 1995 Canadian survey, the point-prevalence for LBP was 29%, and 11% of the population had LBP associated with functional limitation.[9] LBP is the 7th leading reason for visiting a physician (accounting for 2.3% of all visits), the main reason for visiting a neurosurgeon (35%), and accounts for a large proportion of orthopedic surgeons' caseloads (11%-40%) in the USA.[7,10,17-19] LBP is also a recurrent problem, with 60% to 85% of persons with LBP having a recurrence within two years following a first episode.[13,20-22]

However, LBP is, by definition, a clinical symptom or complaint and not an etiologic entity.[11,13,17,23,24] In practice, less than 3% to 20% of conditions generating LBP are given specific diagnoses.[15,25,26] A "non-specific LBP" diagnosis is used in most other cases when the exact pathological process is not identified.[15,17,25,26] Examples of specific diagnoses include observable lesions of vertebrae and discs due to inflammatory diseases, vertebral fractures, and cancer, and LBP from identified visceral causes.[3,26] In contrast, non-specific diagnoses include "back pain," "lumbago," and "sprain or strain" when no attempt is made to confirm the specific source of the pain.[17,26] The majority (80% to 97%) of LBP cases are classified as "non-specific," illustrating that controversy still exists about the source of LBP.

Non-specificity leads to heterogeneous samples in regards to etiology, restricts the development of specific interventions, and creates confusion in research on LBP. Additional research is needed to address questions of etiology and outline factors that may help classify subjects with non-specific LBP into meaningful sub-groups to improve treatment and prevention. This need was identified as the top research priority and primary care challenge at several Primary Care Research Forums on LBP and by leading researchers.[17,27-29] Improving the measurement of disc contour abnormalities may help the research in this priority area because the intervertebral disc is a prime suspect as a source of non-specific LBP. The different etiology hypotheses are discussed in the following sections.

ETIOLOGY HYPOTHESES

Many structures, alone or in combination, have been suggested as possible sources of pain in subjects diagnosed with "non-specific LBP."[30] In theory, any structure in the lumbar area that contains pain-sensitive nerve endings can be a source of pain.[30] Pain-sensitive anatomical structures in the lumbar spine include the outer annulus of the intervertebral disc, facet joint capsules, sacro-illiac joints, vertebral bodies, dura mater, nerve root sleeves, blood vessels, muscles, interspinous, and longitudinal ligaments.[30-32] The disc is the focus of numerous diagnostic and surgical procedures, and is a prime suspect amongst causes of non-specific LBP. LBP etiology hypotheses involving lumbar anatomical structures will be reviewed with a special emphasis on the intervertebral disc and especially on disc contour abnormalities. A short description of the intervertebral disc's anatomy and pathophysiology is presented in Appendix A to support the discussion in the sections to follow.

The Intervertebral Disc as a Primary Suspect for LBP

The disc, a prime suspect in the etiology of LBP, can cause LBP either directly or by influencing nerve roots, or other lumbar structures[3,32,33], and frequently exhibits signs of degeneration. Invasive surgical and discographic studies have proven the intervertebral disc to be a direct source of low back and leg pain.[18,24,33-42] Kuslich et al. found that more than 70% of 193 patients operated on for nerve root decompression reported pain in the back and buttock when the annulus received painful stimuli, and concluded that the intervertebral disc was 'the site' producing LBP.[32] In discographic studies, internally disrupted discs were also found to cause pain in the leg in addition to producing LBP even though positive pain responses were also observed in subjects without symptoms of LBP but with abnormal psychological profiles.[18,39-41,43,44]

In the late eighties, Dr. Vert Mooney, President of the International Society for the Study of the Lumbar Spine, concluded that the answer to the question "Where is lumbar pain coming from?" was "Mainly from the disc."[45,46] Deyo and Phillips reported later that patients with disc-related diagnoses constituted more than half of the LBP caseload of neurosurgeons and about 40% of that of orthopedic surgeons'.[17] More recently, other researchers reported that diagnoses related to the disc account for as much as 39 to 75% of chronic LBP.[8,33,39,40,47] The disc is now the focus of numerous clinical and imaging diagnostic tests as well as of conservative, surgical, and developing biological/genetic therapies.

THE INTERVERTEBRAL DISC AS A SOURCE OF LBP: MECHANISMS

A number of pathological mechanisms, acting alone or in combination, have been suggested to explain how the intervertebral disc can cause low back and leg symptoms. These mechanisms can be classified into mechanical, inflammatory, immunological, and chemical mechanisms.

Mechanical: Following Mixter and Barr's 1934 study, the mechanical mechanism was considered for many years to be the sole cause of discogenic pain.[1,48] Disc contour abnormalities can act as space-occupying lesions that impinge on axons, or the dura, or that compromise the axons' blood supply to create pain.[24,48-52] Differences in the morphology of the disc's outer contour are expected when a mechanical mechanism is at play. Alternatively, pressure applied directly to the outer annulus often produces LBP during surgery under local anesthesia.[32,42,53] In addition, excessive tensile stresses imposed on the annulus' innervated region by injury, degeneration, or disease can also create pain.[24,30,54]

Inflammatory: Markers of inflammation were found in surgical samples from the disc material of persons with LBP but not in healthy discs, and increased concentrations were observed with larger disc contour abnormalities, although the differences were not always significant.[55-63] Macrophages and chondrocytes in degenerated discs release inflammatory cytokines suspected to directly stimulate nociceptive nerve endings, induce the production of nociceptive molecules, or induce further degeneration of the disc matrix.[30,55,62-69] The inflammatory process was also shown to contribute to the resorption of the herniated disc material, especially when prominently exposed to the epidural space.[48,57,70-72] In some studies, the reduction in the size of the disc contour abnormality with conservative treatment was associated with a corresponding change in symptoms [70,72,73] but not in others.[74] The progressive resorption of larger non-contained herniated discs and the associated decrease in pain over time suggests that the size of disc contour abnormalities may also be related to inflammatory symptoms. Alternatively, the inflammation of nerve roots can be triggered by exposure to exudates from the disc.[24,49,75]

Immunological: An auto-immune reaction can occur when the nucleus is exposed to the circulation for the first time through an annulus fissure and is recognized as a foreign tissue by the immulogical system.[76-79] The importance of the immunocompetent response, systemically or at the site of the disc contour abnormality, was found to correlate with the symptoms' duration and was more obvious in patients with more severe disc contour abnormality.[77-80] These findings suggest that the immunological reaction may be proportional to the amount of nucleus material exposed. In animal studies, exposure to nucleus pulposus material led to an elevated concentration of immunocompetent cells at sites where the nucleus was implanted compared to the controls.[49,75,81,82] Furthermore, nerve injury was shown to result from exposure to degenerated or healthy nuclear material but not to the annulus tissue.[83] However, the identification of auto-immune cell clones that proliferate or exhibit cytotoxicity specifically on interaction with autologous antigenic material from the nucleus appears to be difficult, as many researchers have tried and failed to do so.[57,61,84]

Chemical: With disc disruption, the nerve endings in the annulus fibrosus, nerve roots, dura, or dorsal root ganglion can be exposed to enzymes and break down products involved in the disc's degradation process.[30,49,67,79,83,85-87] The pain arising from such exposure is referred to as "chemical nociception." Furthermore, discal chondrocytes are known to produce nociceptive triggers possibly independently from the inflammatory reaction.[49,55,58,66,88,89] Nerve endings, present in the nucleus in 20% of degenerated discs, have been shown to contain and react to nociceptive neurotransmitters.[66,88,90]

In all the discogenic pain mechanisms reviewed here, the amount of disc material exposed beyond the intervertebral disc space boundaries may be a factor influencing the symptomatic or physiologic responses. Considering that chemical irritation and immunological response could lead to inflammation and that inflammation influences the size of observed disc contour abnormalities, an adequate and responsive measure to quantify the degree of disc contour abnormalities needs to be developed. Studies of the relative proportion of disc problems due to mechanical, inflammatory, immunological, chemical or interactions between mechanisms are lacking and could benefit from a quantitative measure of disc contour abnormality.

The Alternative Suspects - Lumbar Structures Causing LBP?

While a prime suspect, the intervertebral disc is not the only lumbar structure suspected of creating LBP. The evidence related to other pain-sensitive anatomical structures is briefly reviewed here to provide a balanced perspective and illustrate how the disc is suspected of being involved in producing pain from some other structures.

In Kuslich et al.'s study described earlier, vertebrae, facet joints, ligaments, and muscles were rarely or never found to be painful even though these structures receive pain-sensitive innervation.[32] In fact, most painful disorders of the lumbar vertebrae have very low prevalence and are considered specific LBP diagnoses because they are relatively easy to diagnose with imaging.[24,30] Facet joints have been shown to be a possible source of pain mainly to the buttock or thigh.[24,30,45,91] In persons with chronic LBP, who were treated at tertiary care centers offering anesthetic blocks, the prevalence of facet injuries, after excluding subjects with radicular symptoms and with imaging findings correlating with symptoms, varied between 15% and 45%, with the highest prevalence observed in older adults.[91-94] It has been suggested that facet degeneration and its related pain could develop secondarily to abnormal facet loading as a result of disc degeneration.[95] Most ligaments are not suspected of being primary sources of LBP because ligaments typically have better elasticity than surrounding tissues.[30] Not surprisingly, a low prevalence (4%) of painful interspinous ligaments has been reported following uncontrolled anesthetic blocks.[30] Stimulations of interspinous ligaments were rarely painful.[53] In contrast, the ilio-lumbar ligament is susceptible to injury but has not yet been shown to receive pain-sensitive innervations.[30,96]

Sprains, spasms, trigger points, and imbalance could affect lumbar muscles.[24] However, no direct evidence suggests that sprained muscles are responsible for LBP.[24,30] The evidence from studies using electromyography and magnetic resonance spectroscopy is inconclusive.[24,97] Depending on the diagnostic criteria used, the prevalence of trigger points varied between 10% and 36% in persons with LBP compared to 5% in control subjects.[98] Even though motor control imbalances have been detected between muscles of the lumbar spine, how an imbalance would lead to LBP is not clear.[99-101] Moreover, the question remains whether muscle disorders precede or follow the appearance of LBP. Irritation of the sacro-iliac joints has also been shown to produce pain over the joint and in the lower limbs.[24,30,102,103] However, in a select group of chronic LBP patients (i.e., in those with negative facet or epidural block, and excluding patients with abnormal lumbar imaging), the prevalence of sacro-iliac problems remained low (15%-19%).[102,103]

The dura mater has been shown to produce pain mostly in the back and the buttock.[30,49,104] Dura mater pain, like nerve root pain, can result from chemical irritation or inflammation occurring when the dura mater is exposed to disc material or when mechanically irritated via the tethering of adhesions developed following severe disc contour abnormalities.[30] Because dura mater pain is often difficult to distinguish from radicular pain, the former's specific prevalence is unknown.[30] LBP can also appear if epidural veins become distended when the blood flow is impaired.[105-107] However, this condition appears to be rare and only secondary to changes in disc contour abnormalities or canal dimensions.[106,107]

While many anatomical structures other than the intervertebral disc have the potential to produce LBP, such instances appear relatively less prevalent. However, the specific prevalence for the different lumbar structures is unclear because the prevalence of the involvement of all suspected lumbar anatomical structures in causing LBP has not been compared within a single study of patients in a primary care setting. Instead, when the prevalence of a given structure's involvement was reported, the samples were often selected to exclude subjects with the involvement of other structures. Nevertheless, disc contour abnormalities and disc degeneration are quite prevalent and suspected to be contributing factors in pain arising from nerve root, dura mater, facet joints, and epidural veins. Therefore, disc involvement is still the prime suspect in most cases of non-specific LBP.

IMAGING OF DEGENERATIVE CHANGES IN THE LUMBAR SPINE

The imaging methods used to assess the condition of the disc are reviewed in the following section, beginning with a historical overview, continuing with MRI, and then discussing the association between imaging findings and symptoms.

Brief Historical Overview of Disc Imaging

Prior to the advent of MRI, the lumbar spine was assessed by using X-rays, but it was not possible to see the disc directly, and only disc space narrowing, not disc contour, could be assessed. X-rays could not be used to detect early disc problems because disc narrowing occurs only when there is severe degeneration.[13,38,108-110] Developments in imaging techniques and contrast material led to the introduction of myelography, discography, and computerized-tomography (CT).[34,35,37] Myelography has the advantage of being the only widely available procedure that allows imaging in standing while weight bearing.[111] However, with myelography, disc contour abnormalities are demonstrated only indirectly by indentation of the distribution contrast material in the spinal canal.[110] Discography has the advantage of determining if the disc is a source of pain because contrast material is injected in awake patients who are asked to grade their symptoms in response to the injection.[44,112] However, these imaging techniques expose subjects to radiation,[36,113] and their invasive nature is associated with a risk for complications such as allergic reactions to the contrast material, infections, headaches, and, possibly, de novo disc contour abnormalities.[114-117] Ultrasound imaging, although often used for non-invasive imaging of soft-tissue, has rarely been used to study the intervertebral disc. This is because finding a location to position the probe and obtain clear images is difficult, especially in narrowed discs.[118-120]

Magnetic Resonance Imaging

MRI was introduced relatively recently and is believed to be free of risk for appropriately screened subjects, allowing for the study of asymptomatic subjects or repeated imaging in follow-up studies.[14,16,121-126] MRI provides a much better resolution for soft tissues than imaging using X-rays.[16,124-130] Modic et al., in fact, concluded that MRI was better than any other technique for studying disc contour abnormalities.[110]

The MRI signal is a function of a tissue's water content.[121,124,125] MR creates images based on the relative concentration of hydrogen ions in the areas being imaged instead of their opacity to X-rays, leading to better soft tissue resolution. Different contrasts can be selected with MRI to best reveal specific degenerative findings.[129,131,132] For disc contour abnormalities, proton density images are helpful in judging annulus integrity, and T2-weighted images help to reveal the posterior longitudinal ligament line and the signal intensity of the nucleus.[129] In addition, sagittal images can be obtained more easily with MR than with CT and sagittal images are useful for demonstrating disc abnormalities.[111,133] Sagittal images can determine whether the disc content from contour abnormalities has migrated caudally or cranially, as it commonly does in these cases.[134] Thus, MR is the imaging procedure of choice for case-control and longitudinal studies of the association between degenerative disc changes and their determinants or symptoms. However, because of the limited availability and the cost of MRI examinations, such studies could not be widely conducted.

Qualitative Findings on Lumbar Spine MR images

While MRI is the method of choice to study disc contour abnormalities such as bulges and herniation, other disc-related findings are also recorded on MR images. The following paragraphs briefly discuss a variety of disc-related findings that can be observed on MR images of the lumbar spine (see also Appendix B for definitions). The controversies surrounding the clinical relevance of the imaging findings most related to disc degeneration are summarized in the following section when discussing the prevalence of the disc-related findings in subjects with and without symptoms. Disc contour abnormalities are associated with a number of imaging findings presented below.

Osteophytes: Osteophytes are defined as "focal hypertrophy of bone surface and/or ossification of soft tissue attachments to the bone."[135] Osteophytes are thought to arise from new bone created in response to traction on the bony attachment of Sharpey fibers when disc contour abnormalities of a degenerated disc form.[13,136,137] In an autopsy study, vertebral rim osteophytes were noted in all cases with degenerated discs[137] and were also said to be a definite sign of chronic disc contour abnormality.[111]

Disc Signal Intensity: Because of its higher hydration, the normal nucleus material has a high intensity signal on T2-weighted images while the annulus appears dark.[110,121,130] Decreases in signal intensity indicate a decrease in water and glycosaminoglycans' content which is usually concomitant with an increase in collagen content as a result of the degenerative process. [38,109,130,138-141]

Annular Tears - High Intensity Zone (HIZ): Annular tears detected by discography are a good predictor of concordant pain reproduction.[37,112] However, annular tears are not observed directly on MR images. Instead, a HIZ, an area of high signal intensity within the outer annulus of the disc on T2-weighted MR images, is thought to reflect the presence of a symptomatic tear of the annulus.[38,113,135,142] However, controversy still exists over whether the HIZ is a sensitive and reliable indicator of symptomatic annular tears.[18,38,43,113,115,135,142-144] MRI-discography, although more invasive than MRI without injection of contrast material, appears promising as an alternative to detect the presence of annular tears.[117]

Disc Height Reduction: As with X-rays, the change in disc height, the distance between the endplates over and under the disc, is often used with MRI as a general indicator of disc degeneration alone or as part of global degeneration scores.[127,135,145-147] A relationship between disc height and disc contour abnormalities or loss of nucleus material has been described.[38,148-150]

Stenosis: Stenosis can be assessed by judging the spinal canal area's degree of compromise on axial images.[135] Stenosis has also been evaluated as follows: reduction the of the vertebral canal's anteroposterior diameter larger than seven millimeters,[128] as the presence of areas of anterior and posterior encroachment of the sub-arachnoid space,[130] and as the obliteration of the epidural fat with flattening of the thecal sac.[151] A stenosis rating scale considering the encroachment on the central canal and the amount of cerebrospinal fluid still present around the nerve roots has also been discussed.[152] Analysis of these definitions suggests that stenosis is another MRI finding that can be measured quantitatively by using geometric measures of the canal, nerve tissue, and disc contours.

Nerve Root Compromise: An ordinal scale for neural compromise was developed by Boos et al. based on the relationship between the disc material and the nerve roots, the thecal sac, or the epidural fat.[122,153] This rating scale correlated well with surgical findings.[153] Other researchers also recommended specifying if the compromise involves the lateral recess, the intervertebral foramen, or the spinal canal.[128]

Schmorl's Nodes: Schmorl's nodes findings can be considered intravertebral contour abnormalities.[135] They are a localized displacement of nucleus, cartilage, fragmented apophyseal bone, or fragmented annular tissue beyond the intervertebral disc space.[111,135] Schmorl's nodes are seen as small indentations of the low-signal intensity sub-chondral bone of the endplate on T1-weighted images, which contain high-signal-intensity disc material on T2-weighted scans.[154]

Bulging: The definition of "bulging" proposed by the Consensus of the leading North American Radiological Societies is presented in Appendix B.[135] Bulging is often rated as present or absent,[146] graded from mild to severe,[155] or as a mild finding in qualitative scales assessing the severity of disc contour abnormality (e.g., normal, bulge, protrusion, extrusion, and sequestration).[156-160] Some authors have suggested that bulging could indicate the following differential diagnoses: a normal anatomic variant, an illusion caused by the volume-averaging effect, normal aging remodeling, pathologic disc degeneration, or posterior disc rupture with sub-ligamentous herniation.[111,161-164] However, in autopsy and discography studies, annular tears were found in most bulging discs.[165]

Herniation: The Combined Task Force on the classification of lumbar disc pathology from the leading North American Radiological Societies proposed a clear definition of "herniation" (Appendix B).[135] However, a wide variation of herniation definitions is reported in the literature.[14,38,128,135,146,151,157,160,161,166,167] Some authors differentiate among bulging, protrusion, extrusion, and sequestrated findings, which they consider a continuum of progressively more severe disc contour abnormalities.[156-160] Others consider that various combinations of the findings above are manifestions of the same finding: herniation.[4,14,123,168] Although, researchers recommend determining if a herniation is contained or not, MRI's technical limitations usually preclude distinguishing between a contained and an uncontained disc herniation.[135] Researchers also recommend that the size and location of herniation be reported relative to the size of the spinal canal, but most qualitative scales do not assess the magnitude of the herniation.[161] Further small, medium, and large qualifiers are considered unreliable in quantifying the size of herniations.[161]

In light of this review, the development of a quantitative measure of the amount of disc material located beyond disc space boundaries, of nerve root compromise, and of spinal canal dimensions would address many of the recommendations made for the measurement of disc contour abnormalities, stenosis, and nerve compromise for clinical purposes. Ideally, such a measure should also consider the association between osteophytes and disc contour measurements.

ARE DEGENERATION FINDINGS ASSOCIATED WITH LBP SYMPTOMS?

With MRI, more asymptomatic subjects have been evaluated. Large variations in the prevalence of the imaging findings have been observed in both symptomatic and asymptomatic individuals. This wide variation raised questions about the meaningfulness of many MRI findings.[121,127] The relationships between anatomic variations and low back symptoms are still speculative and controversial, and better measures of anatomic impairments are needed for further studies.[40,47,121,122,127,138,150,151] To help determine whether various disc-related MRI findings are specific to patients with LBP, a review was conducted focusing on studies of the prevalence of different imaging findings (1) in samples of subjects without symptoms (Table 1.1), (2) in subjects with LBP (Table 1.2), and (3) in population samples or occupational group samples selected without regard to LBP (Table 1.3).

The range of reported prevalences of degenerative findings for asymptomatic subjects was 10% to 81% for bulging, 3% to 63% for protrusion, 0% to 24% for extrusion, 6% to 83% for reduction in signal intensity, 3% to 56% for disc narrowing, 6% to 56% for anular tears (HIZ), and 8% to 19% for Schmorl's nodes (Table 1.1). No study reported on the prevalence of vertebral osteophytes, and none found sequestrations in asymptomatic subjects. The prevalence of spinal canal stenosis in persons without LBP was low (0-10%).[127,146,151,152,160,168-170]

The corresponding prevalence in the samples of symptomatic subjects reported in the studies reviewed were 37% to 90% for bulging, 2% to 54% for protrusion, 35% for extrusion, 13% to 57% for reduction in signal intensity, 3% to 41% for disc narrowing, 3% to 60% for anular tears (HIZ), and 6% to 23% for Schmorl's nodes (Table 1.2). No studies reviewed reported the prevalence of vertebral rim osteophytes. Sequestrations were observed in 7% of symptomatic subjects,[122] and canal stenosis was observed in 10% to 50% of subjects with LBP depending on symptoms' severity.[168]

The corresponding prevalence in the samples of subjects not selected on the basis of the absence of LBP (i.e., population-based or occupational group samples) were as follows: 22% to 50% for bulging, 0% to 70% for protrusion, 0% to 55% for extrusion, 0% to 0.2% for sequestration, 9% to 86% for reduction in signal intensity, 13% to 53% for disc narrowing, 7% to 40% for anular tears (HIZ), 6% to 79% for Schmorl's nodes, and 6% to 57% for vertebral rim osteophytes (Table 1.3). Central canal stenosis was reported in one study to be present in 12% of 40 year old subjects, and moderate to severe dural sac compression was observed in 9% to 44% of disc levels in adults.[150]

Severe disc contour abnormalities (extrusion, sequestration) appeared to be slightly more prevalent in patients with symptoms. However, given the wide ranges of prevalences reported for all findings, no clear difference in the prevalence of degenerative disc findings was found among 'asymptomatic samples', 'symptomatic samples,' and 'population-based or occupational group samples.' The differences among the studies in regards to the subjects' age, exposure to known and unknown risk factors, and the disc levels included, may have contributed to the variations reported in the prevalence rates (Table 1.1 and 1.3). However, for findings such as lower grades of disc contour abnormalities (bulging and protrusions), the differences in the reported prevalence were as high as 40% within a given age group, suggesting that age alone does not explain the differences.

The differences in the operational definitions of the imaging findings may also account for much of the variation. The variations in the definitions used for the assessment of disc contour abnormalities are particularly important with authors rating individually or in combination bulging, hernation, protrusion, extrusion, sequestration, prolapse, contained, and non-contained. Brant-Zawadski et al. showed that calculated prevalences vary depending on the operational definition used to rate disc contour abnormality.[157] Future research would benefit from the use of standardized definitions of imaging findings[171] such as quantitative measures of disc abnormality.

Differences in Prevalence Between Subjects Matched for Confounding Factors: Although asymptomatic subjects had more abnormal findings than expected, in studies where symptomatic and asymptomatic subjects were matched for suspected counfounding factors, subjects with symptoms had a higher prevalence of severe disc extrusions (21-35% vs 3-13%), severe stenosis (50% severe LBP, 11% moderate LBP, 3% asymptomatic), and major neural compromise. (54% vs 4%),[122,168,169] but not general disc degeneration (reduced signal intensity).[14,16,122,172,173]

When comparing the prevalence of the findings for asymptomatic subjects with different histories of LBP, the prevalence of general disc degeneration, disc narrowing, low grade of disc contour abnormalities, annular tears, or Schorml's nodes did not differ significantly.[125,152,174,175] The prevalence of severe contour abnormalities or nerve compromise was usually too small in some of the asymptomatic groups with different LBP histories to allow for valid testing of the statistical differences between groups.[125,152,174] However, the odds ratios for a more frequent past history of LBP in asymptomatic subjects at the time of examination were significantly larger for subjects with severe disc contour abnormalities (OR 32.54 vs Never LBP)[152] or with moderate or severe stenosis (OR 5.6 vs never LBP).[152,174] Extrusion and nerve compression findings suggest that a space-occupying mechanism is at play. Quantitative measures may help clarify the associations between these findings and symptoms.

Correlations between Changes in MRI Findings and Pain in Follow-up Studies: Longitudinal studies of the relationship between progression in imaging abnormalities and symptoms can demonstrate that the suspected causal change occurred within the period when the pain developed. In adults, follow-up studies assessing general disc degeneration showed worsening over time, but that worsening was not associated with the appearance of symptoms.[125,160,167,176] The degree of disc contour abnormality did not worsen significantly in 3- to 5-year follow-up studies, and changes were not associated with the appearance of symptoms.[167,169,176] However, in those studies, only between 11% and 33% of the initially asymptomatic subjects (n=31 to 46) developed LBP symptoms interfering with activities (work or leisure) for more than a day, thereby limiting the power of the statistical analyses.[167,169,176] Nevertheless, Boos et al.'s 5-year follow-up study showed that the development of neural compromise had an impact on LBP-related medical consultations and work absence.[167] In a 7-year follow-up study of 31 initially asymptomatic subjects, Borenstein et al. showed that herniation, bulging, stenosis, and global degeneration all had worsened.[160] Furthermore, the relative risk that LBP would develop in individuals with worsening abnormalities on MR images was 3.5.[160] Sensitive quantitative measures of disc findings would be necessary to determine whether changes in disc degeneration during relatively short follow-up periods were associated with symptoms, especially if larger samples are not feasible.

ASSESSING DISC DEGENERATION ON MR IMAGES – NOT ONLY TO SEARCH FOR A CAUSE OF LBP

The interest in disc contour abnormalities has arisen not only because the disc is a prime suspect as a source of LBP but also because disc degeneration is now the focus of intensive research. During the last two decades, technological developments have made possible the study of the biological and genetic bases of disc degeneration.[177] Interest in the development of biological and genetic therapies to prevent and reverse disc degeneration is steadily growing.[177] Twenty-eight percent of the podium research presentations at the 2005 meeting of the International Society for the Study of the Lumbar Spine were related to biological and genetic disc research, attesting to this increased interest. The interest in such therapies has arisen because disc degeneration can be a source of LBP and is believed to be the start of a degenerative process which in turn leads to different spinal disorders.[178] The assessment of the interrelation between different degenerative signs of the lumbar spine would also benefit from valid, reliable, and sensitive measures of disc degeneration, including measures of disc contour abnormalities.

SUBOPTIMAL MEASUREMENT PROPERTIES OF CLINICAL RATINGS

Assessments of disc contour abnormality from MR images are often used to guide surgical decisions when correlating with clinical findings. MRI measures are also considered in correlation studies of the determinants and effects of disc contour abnormalities. However, for clinical decisions on an individual basis, measures should demonstrate high reliability and be valid. Reliability coefficients between 0.85 and 0.94 are considered a minimum by some authors for inference at the individual level.[179] In research, where group inferences are of interest, reliability coefficients over 0.70 are usually considered a minimum.[180] High reliability coefficients are also required to avoid attenuation of correlation estimates.[179,181] Previous reports of the reliability of imaging findings' ratings show that the reliability of qualitative ratings of disc degeneration and, especially, disc contour abnormality is suboptimal. Reliability estimates vary with rater training or speciality,[155] the scale used,[38,152,157-159,167,168] the disc level considered (reliability is usually poorer in the lower spine),[122,155,158,168,182] and severity (more severe degeneration provide more reliable rating).[38,155,158,182] The literature on reliability and validity evidences for MRI ratings of disc contour abnormalities is reviewed in this section and demonstrates that the psychometric properties of such qualitative measures needs to be improved.

Reliability: For ordinal rating of lumbar disc herniation (combinations of at least protrusion, extrusion, sequestration, but not bulging alone), inter-rater Kappa (κ) coefficient varied between 0.24 and 0.96,[38,122,157-160,168,183] and weighted-κ coefficients varied between 0.50 and 0.75.[38,152,158,167] For intra-rater reliability, κ values varied between 0.65 and 1,[157,159,168] and intraclass correlation coefficients (ICC) varied between 0.58 and 0.80.[184] Brant-Zawadski et al. found that a rater could more easily distinguish between herniation and non-herniation or between normal and abnormal disc contour (κ

= 0.75-0.80) than determine whether the disc was normal, bulging, protruded, or extruded (κ = 0.57-0.59).[157] Boden et al. also reported that raters disagreed more often on the severity of an abnormality than on its presence or absence.[146] This finding demonstrates that for correlation studies, obtaining adequate reliability for a qualitative measure of disc contour abnormalities with multiple increments may be difficult.

The inter-rater reliability of measures of the severity of disc bulging (assessed separately from herniation) is as follows: κ coefficient of 0.36,[160] and ICCs between 0.49 and 0.64.[155,182] For intra-rater reliability, ICCs varied between 0.58 and 0.80.[155,182,184] Jarvik et al. and van Rijn et al. showed that raters had more difficulty reliably rating the lower grades of disc contour abnormalities than the severe ones.[158,183] This observation is supported by the findings reported above for the reliability of bulging comparatively to that of herniation ratings.

The intra- and inter-rater reliability of other disc-related findings observed on MR images, such as neural compromise, spinal stenosis, disc height, and endplate irregularities was similar even though the upper limit of the reliability estimates reported for endplate ratings was lower (intra-rater κ = 0.40-1.00 [128,153,155,159,168] and ICC= 0.54-0.93[155,182,184]; inter-rater κ = 0.27-0.87 [38,128,152,153,155,159,160,167,168,183,185] and ICC = 0.51-0.69[155,182]).

Raininko et al. published a reliability study involving three raters using written instructions and standard images.[155] This study used a sample of symptomatic and asymptomatic subjects, which were used in some of the studies in this dissertation. The mean intra-rater κ coefficient observed for ratings of disc herniation between L4-S1 was 0.70, and the ICC for bulging was 0.76.[155] At L4-S1, the mean intra-rater κ coefficient for the evaluation of osteophytes was 0.65, and the mean intra-rater ICC's for disc height was 0.84.[155] the corresponding inter-rater κ coefficients were lower, varying between 0.30 (disc herniation) and 0.47 (osteophytes).[155] The inter-rater ICC's were also lower and varied between 0.57 (bulging) and 0.61 (disc height).[155] These results illustrate that the reliability of disc contour abnormalities is sub-optimal, and that the inter-rater reliability is inadequate for both group- and individual-level inferences.

Lower levels of reliability estimates can be expected, on average, for subjects in the normal range of degeneration compared to subjects presenting extensive degeneration.[186] This expectation may explain the lower reliability observed in studies by Brant-Zawadzki et al., Jarvik et al. and Raininko et al. where samples included asymptomatic subjects, who would typically not seek imaging of their spines.[152,155,157] Boden et al. also observed better agreement in their group of subjects with symptoms than in their asymptomatic group.[146] These findings suggest that in studies where the prevalence of severe findings is expected to be low (e.g., studies of the determinants of disc degeneration conducted in population-based samples rather than patients with suspected back problems consulting a physician), qualitative ratings of disc findings would likely be inadequate.

Validity: Evidence of the validity of the qualitative assessment of disc morphology is limited.[117,133,153,187-194] This lack of evidence can be attributable to many factors, the most important of which is probably the lack of an appropriate gold standard for criterion validity studies. The validity of qualitative ratings of the presence of disc contour abnormalities on MR images has been assessed by comparison with CT-discography (69-100% accuracy)[38,133,187-189] and surgical findings (77-96% accuracy).[129,187-190,195] Studies of the validity of magnitude ratings for abnormal disc contour are rare. In two series of 13 and 31 discs, the MRI accuracy for the grade of disc contour abnormality assessed at surgery was 38% and 81%, respectively.[129,188] In addition, the qualitative MRI assessment over- and underestimated the grade of contour abnormality found at surgery.[129,188] Validation studies often involved inadequate samples sizes.[129,133,188,196] In many validity studies, not all the discs imaged with MRI had been assessed by the reference modality[188-190] or the validity of the MRI findings was retrospectively examined after the subjects had had surgical exploration.[194,195] The validity estimates may have been positively biased if subjects were more likely to have surgery when imaging suggested severe contour abnormality, or negatively biaised if subjects were more likely to be imaged with a second modality if the initial findings were ambiguous.

The validity of other imaging findings closely related to disc contour abnormalities was assessed in a few studies summarized here. However, when differentiating contained from non-contained disc material, the accuracy of MRI for predicting surgical findings varied widely (42% to 80%).[129,194,196] In contrast, studies of the qualitative assessment of the magnitude of nerve root compression showed that MRI gradings correlated highly with the surgical findings (Spearman r = 0.86, 83-96% accuracy).[153,192]

Discriminative / Longitudinal Validity: The results of some studies suggested that measurements of the extent of disc contour abnormalities with more increments have the potential to help clarify the association of such findings with suspected determinants or symptoms.[122,197] Brant-Zawadski et al. compared two ordinal rating systems for disc contour abnormality and found that the scale with more levels had a higher association with pain.[157] Furthermore, in a 5-year follow-up study, two experienced raters observed no changes in the ordinal grade of disc contour abnormality, while two independent raters using a qualitative rating of change concluded that some degree of worsening had occurred in 25% of the subjects.[167] The inadequacy of the qualitative ratings of disc contour was also illustrated in other studies where authors could not demonstrate changes between repeated measures using traditional ratings (i.e., bulge, protrusion, extrusion, sequestration) rather than ratings of change.[198,199] In fact, in longitudinal studies, new qualitative ratings are commonly developed and used to detect changes in disc contour abnormalities [70,200-205] without assessment of their measurement properties.[200,201,203]

In summary, qualitative measures of disc contour abnormality have suboptimal reliability, particularly when raters are distinguishing between low grades or between multiple subtle differences in the degree of disc contour abnormality. This suboptimal reliability may affect the ability to use qualitative ratings of disc contour abnormality in studies of their determinants and symptoms. While validity evidence exists for the presence or absence of severe disc abnormalities, it was produced in studies of low quality. Furthermore, validity evidence for the presence or absence of smaller grades of disc contour abnormalities and for magnitude gradings is lacking. Finally, the fact that in most studies measuring changes in the magnitude of disc contour abnormalities, different qualitative rating scales are developed, illustrates that a qualitative grading from normal to herniation is inadequate. Thus, valid and reliable measures of the magnitude of disc contour abnormalities are needed. This needs supports value of developing continuous measures of disc morphology for exploring associations with symptoms or studying longitudinal changes.

Quantitative Measures

EXPECTED ADVANTAGES OF QUANTITATIVE MEASURES OVER SUBJECTIVE RATINGS

Quantitative measures of disc degeneration for MR images may help address the reliability limitations outlined in the previous section for qualitative measures. From the review of the prevalence of disc abnormalities in asymptomatic subjects, severe disc abnormalities, canal stenosis, and nerve compressions appear to be more specific to persons with symptoms. However, quantitative measures of disc contour abnormalities have not been studied as extensively as spinal stenosis measures. As well, the existence of an association between nerve root compression and symptoms is not as controversial as for disc contour abnormalities. A valid continuous quantitative measure of disc contour with reliability superior to that of qualitative ratings could help to

- identify the determinants of disc contour abnormalities,
- detect changes in the size of disc contour abnormalities as a result of natural history, biological, conservative, and surgical interventions,
- study the effect of different postures and loading conditions on disc morphology,
- determine the prognostic value, if any, of the extent of disc contour abnormalities on conservative and surgical outcomes, and
- clarify the role of disc contour abnormalities in LBP causation.

The areas listed above correspond closely with the research priorities developed at a 1995 workshop entitled "New Horizons in Low Back Pain" which was supported by the American Academy of Orthopedic Surgeons, the National Institute of Arthritis and Musculoskeletal and Skin Disorders (NIAMS), and the National Institute of Health (NIH).[206] These research priorities are outlined below.

- Study the accuracy, reproducibility, and validity of the parameters used in the diagnosis of herniated discs including imaging studies.[p156]
- Improve measurement methods to identify the risk factors for disc degeneration and herniation, and the approximate magnitude of their effect.[p157-8]

- Study the natural history of disc abnormalities seen on advanced imaging studies.[p155 and 751] Longitudinal studies are needed to identify persons with and without specific anatomic lesions, and to follow them at intervals for development or resolution of anatomic lesions and/or symptoms.
- Establish the therapeutic utility (e.g., reliability, impact on care) of diagnostic tests commonly used in the investigation of idiopathic LBP with a high priority for MRI. [p469]
- Identify factors determining when and whether disc structural changes induce LBP.[p628]

Upon reviewing these priorities, it is evident that attempting to improve the measurement of disc contour abnormalities appears justified and would likely be quite useful in research.

For researchers, the main advantage of a continuous quantitative measure of disc contour would be to improve the ability to study the effect of risk factors on disc contour abnormalities.[13] With quantitative measures, inter-evaluator error, and test-retest variability due to subjective factors affecting the evaluator would likely be reduced. In addition, continuous quantitative measures of disc contour abnormalities would be less likely to exhibit restriction of the range, and thus, estimates of the correlations between risk factors or pain and degeneration would not be as attenuated due to the restriction of range as they are with qualitative ratings. Many authors have now demonstrated by using a wide variety of measures that, in most cases, the size of disc contour abnormalities decreases over time occasionally with a corresponding improvement in symptoms.[70,72,202-205,207-209] The examination of such an association would also benefit from a reliable quantitative measure of disc contour abnormalities. Adequate documentation of the natural history of imaging findings in persons with and without LBP would also be useful for interpreting the results of clinical trials when one arm of the study is non-interventional.[210] A standardized quantitative measurement of disc contour abnormalities, if found to be of consequence, could help obtain uniformity in sub-classifying persons with non-specific LBP. Such uniformity is essential to compare research results regarding the diagnosis and treatment of disc disorders related to disc contour abnormality.[111]

Even though not all hypotheses implicating the disc in the etiology of LBP suggest a direct association with contour abnormality, the influence of degeneration on the morphology of the disc is still considered a potential cause of LBP. Also, in inflammatory, immunologic, or chemical discogenic pain mechanisms, a relationship, albeit possibly non-linear, may exist with the amount of disc material located beyond disc space boundaries. This possibility may be better studied with a quantitative measure of disc contour abnormalities than with a 4-point qualitative rating.

QUANTITATIVE MEASURES: WHAT PROGRESS HAS BEEN MADE SO FAR?

The literature on quantitative measures of disc degeneration will be reviewed here to illustrate their potential to improve reliability and to provide a tool in research on the determinants of disc degeneration and sources of LBP symptoms. This literature review also provides a historical perspective on the development of quantitative measures of disc-related degenerative findings on MR images.

Signal Intensity

Signal intensity was one of the first MRI findings involving the disc to be studied quantitatively. At first, most clinicians and early investigators subjectively rated signal intensity by using ordinal scales from printed films.[211] Reductions in signal intensity were associated with symptoms or with other indicators of degeneration in such studies.[38,138,141,212,213] These observations led to increasing interest in signal intensity. In the last fifteen years, some researchers have suggested studying signal intensity quantitatively directly from the digital image data.[139,214,215]

To address the effect of spatial variations in coil sensitivity on signal intensity measurements, researchers have normalized the disc signal intensity relative to that of the cerebrospinal fluid.[214,216,217] Videman et al. concluded: "Based on findings from both living subjects and cadaver specimens, digital assessments of signal intensity offer more precise and detailed information about general disc desiccation than do subjective readings."[214] The digital measure of signal intensity had better intra-observer reliability and correlated better with potential determinants of LBP than a qualitative rating of signal intensity.[182] For example, a significant association between the quantitative measure of disc signal intensity and two polymorphisms of the vitamin D receptor gene was revealed when using the quantitative measure but was undetected when using a less reliable qualitative grading scale.[182,214,218] High estimates of intra-evaluator reliability for quantitative signal intensity measurements have been provided (ICC = 0.95-0.98[182],0.77-0.97[139]; r = 0.93-0.99[214]). The observed advantages of a continuous digital measure of signal intensity over the traditional qualitative ratings suggest that quantitative measures for other disc abnormalities could also outperform their traditional qualitative rating counterparts.

Disc Height

Before MRI evaluation became available, researchers were already interested in continuous measures of disc height, and suggested numerous methods to measure disc height from X-ray films.[148,219-229] Quantitative measures of disc height from MR images were also developed.[139,182,230-235] Some authors simply measured the shortest distance between the anterior and posterior edges of the neighbouring endplates.[139,230,232] The reliability estimates of the quantitative measures of disc height were variable: intra- and inter-rater ICCs between 0.57 and 0.95,[139,182,232] and measurement error between 0.2 mm and 0.5 mm, depending on the method used.[230,232-234] Battie et al. averaged the distances between the anterior and posterior corners of the adjacent vertebra, and a distance between the endplates anteriorly and posteriorly to the nucleus.[182] However, the quantitative measures were not more reliable than the qualitative ratings, and the authors attributed this finding to difficulties in identifying bone and disc boundaries in the periphery where intensity is often low on MR images.[182] Edmonston et al. were able to address this difficulty by measuring disc height only near the nucleus and obtained better reliability (ICC= 0.94).[234] Such difficulties should be considered when developing measures of disc contour abnormality if the vertebral corners are used as a reference from which to measure.

Whole Disc Morphology

Some authors have measured disc diameters to quantify disc contour abnormalities and found the effect of diurnal variation or of different low back postures to be minimal.[236,237] Any changes using such measurements, if present, were likely too small relative to the overall disc diameters to be detected. Other authors measured the total disc volume by adding the product of the cross-section area outlined on each MRI slice where the disc could be seen by the slice thickness and demonstrated small diurnal variation.[237-240] The error associated with repeated measures of disc diameters or volumes was small (< 5%), and the differences between repeated measures were not significant.[237-240] However, disc volume and diameters are indicators of general disc morphology and are not specific to disc contour abnormalties.

Disc Contour Abnormalities - Any Plane

The volume of disc contour abnormalities, the sum of the product of the axial or sagittal areas of contour abnormality by slice thickness, has also been calculated.[208,241,242] However, the reliability of this time-consuming procedure, requiring segmentations of many MR images, has not been studied adequately.

Disc Contour Abnormalities – Axial Plane

Quantitative measures of disc contour abnormalities, using axial MR or CT images, have been suggested. Ratios of the disc material's area located beyond the disc space boundaries to the spinal canal's area have been used in research.[74,197,241,243-250] Ratios of the linear dimension of disc contour abnormalities and spinal canal dimensions were also computed for findings on MR and CT images.[73,197,209,247,248] Reports of reliability for such measures, although sparse, are promising. The inter-rater error was estimated to vary between 2.1% and 5% for the average of three measurements.[197,244,245] Alternatively, some authors have simply measured on axial images the longest distance between the apex of a disc contour abnormality within the spinal canal and the estimated position of the normal disc boundary.[251] The reliability of such distance was not assessed.

Validity evidence for the quantitative measures described above, as with most other measures of disc morphology, is rare. The main difficulty with axial measures of disc contour abnormality is finding a reference point (anatomical landmark such as vertebral bone or normal annulus position) from which to measure the extent of the contour abnormality. Insufficient details on the segmentation procedures used for the axial measures proposed to date limit the ability to replicate such studies. Nevertheless, convergent validity evidence was provided in a study where three ratios of disc abnormality to spinal canal area were highly correlated with each other ($r > 0.73$-0.98).[248] Moreover, using a threshold of 6.8mm for disc contour abnormalities extending into the spinal canal has a 98% positive and a 69% negative predictive value for the presence of symptoms.[247] Ratio indexes demonstrated some discriminative validity in distinguishing subjects with and without pain.[247,252]

The findings reported above show that precise measurements of disc contour abnormalities on MR images are potentially useful. Axial measures demonstrated promising reliability but their interpretability may be limited by the lack of standardization in identifying a reference point from which to measure the contour abnormality. Another limitation of axial measures is that disc material located beyond the disc space boundaries reportedly migrates caudally or cephally in 65% to 79% of cases.[134,253,254] Thus, axial measures obtained at mid-disc level are most likely not adequate, alone, for thoroughly quantifying the size of migrated disc material.

DISC CONTOUR ABNORMALITIES – SAGITTAL PLANE

In the sagittal plane, distance measures of disc contour abnormalities[182,255-259] and ratios of the dimension of the disc contour abnormality to that of the spinal canal[209,241,243,260] have been obtained from CT and MR images. However, the descriptions are usually insufficient to allow replication, and reports on reliability are typically questionable due to methodological problems (i.e., sampling,[256,257] statistics used,[256,259] and dichotomizing of the measure[255,261,262]). The reliability estimates reported for such quantitative measures of contour abnormalities in the sagittal plane vary widely (ICC = 0.40 to 0.97,[182,257,261] standard deviation of root mean squares of repeated intra-rater measurements 0.29mm,[259] inter-evaluator and inter-test Spearman's rho= 0.96-0.98[256]; and limits of agreement = 0.2 to 0.9 mm[256]). Validity evidence of the sagittal quantitative measures of disc contour abnormalites is quite rare. Cooley et al. showed that measures of the distance of disc contour abnormalities, using a ruler, correlated strongly with measures obtained with a digitizer (rho > 0.97).[257] Moreover, the mean measures obtained with those two methods were significantly different for subjects with normal, bulging, protruding, or extruding discs assessed qualitatively.[257] Limited evidence is available of the ability of quantitative measures of disc contour abnormalities to detect small differences such as changes in size over time[209,243] or diurnal variations in disc contour, while such differences were undetected when using qualitative ratings.[263]

The methodological problems with the reliability studies and the wide variability in reliability estimates reported to date suggest that further development and assessment are needed for quantitative measures of disc contour abnormalities from sagittal images. Furthermore, with distance measures of how far a disc contour abnormality extends into the spinal canal obtained only at mid-disc height or within the intervertebral space, disc material could be present above or below the point where the distance is measured and not be detected. This possibility is of concern because researchers have reported that annular tears were more frequent at the insertion of the disc in the vertebral body than at mid-height and that disc material often migrated caudally.[134,264] To my knowledge, only one other study of quantitative measures of disc contour abnormality has used the area in the sagittal plane as part of a volume calculation. Further assessment of the measurement properties of quantifying the area of disc contour abnormalities on sagittal images is needed.

In further developing quantitative measures of disc contour abnormalities on sagittal MR images, the presence of osteophytes and difficulties in identifying bone and disc boundaries near the vertebral corners need to be considered. Milette insists that: "the peripheral margins of the disc [are] meant to refer to the initial or original limits of the intervertebral space, exclusive of any osteophytes that may have developed."[111] Indeed, the presence of osteophytes complicates the measuring of disc contour abnormalities because with osteophytes, identifying reference points from which to measure is difficult.[155,257,265]

To address this issue, Harada et al. suggested a standardized procedure to identify a position from which to measure the area of disc contour abnormality on sagittal images by using the mid-point of the posterior vertebral wall of the vertebra above and below the disc.[265] To my knowledge, this procedure is the only area that has been recommended for measuring disc contour abnormality, and its reliability has not been assessed adequately. However, this procedure does not consider that, on mid-sagittal MR images, identifying the position of the posterior vertebral wall is difficult because blood vessels are seen entering the vertebral body at its midpoint. Harada et al.'s method may be adapted by using a point not at mid-height, but rather, on the posterior vertebral wall at a position less likely to be influenced by progressive degeneration of the vertebra and where the vertebral artery is not seen. Such a modification should be assessed on mid-sagittal images where the vertebral artery is susceptible to influence the segmentation of the posterior vertebral wall.

THESIS OBJECTIVES

The objective of this thesis was to develop quantitative measures of disc contour abnormalities on MR images, to address the limitations and recommendations outlined in this introduction, and to assess the measures' measurement properties. Four quantitative measures of disc contour abnormalities for MR images were developed. The following chapters include a description of these measures. Each chapter's objectives are listed below.

The primary objective of the study reported in Chapter 2 was to determine, by using generalizability theory, the reliability of the proposed quantitative measures of lumbar disc contour abnormalities. Secondary objectives were to (1) compare the reliability of quantitative measures to that of qualitative measures, (2) compare the reliability of distance and area measures of disc contour abnormalities, and (3) determine the size of error due to image resolution, intra-, and inter-evaluator differences.

In Chapter 3, the reliability of the quantitative measures of disc contour abnormalities is assessed further according to the following objectives: (1) to determine the reliability of quantitative measures of lumbar disc contour abnormalities on MR images obtained on different MR scanners, (2) to determine the proportion of measurement error due to differences between MR scanners (including subject repositioning that occurs in test-retest situations) and between repeated measures from a single evaluator, and (3) to confirm the results from the comparison of the reliability of distance and area measures of disc contour abnormalities.

The primary objective of the study reported in Chapter 4 was to determine the effects of flexing and extending the lumbar spine on the size of disc contour abnormalities assessed quantitatively. More specifically, (1) the size of lumbar disc contour abnormalities measured quantitatively in subjects with the spine in flexed, neutral, and extended positions within the constraints of the MRI equipment, and (2) whether large disc contour abnormalities were associated with larger differences in the dimensions of contour abnormalities between positions were examined.

In Chapter 5, the convergent and divergent validity of the quantitative measures of disc contour abnormalities is assessed to address the following objectives: (1) to investigate convergent and divergent validity evidence to support the use of midsagittal quantitative measures of disc contour abnormalities based on correlations between the quantitative values and qualitative ratings made by medical specialists or amongst different quantitative measures of contour abnormalities, and (2) to compare the dimensions of quantitative disc contour abnormalities obtained from midsagittal or para-sagittal images, and to provide insight on which slices quantitative contour abnormalities should be measured.

Chapter 6 concludes with an integrated discussion. First, a summary of the findings is presented. Then, the strengths and limitations of the research completed in this thesis are discussed. Suggestions for future research to continue the development of the quantitative measures proposed are then provided. Some areas where the quantitative measures proposed could be used are discussed along with potential challenges associated with such projects. Chapter 6 concludes with recommendations resulting from this thesis work.

Table 1.1. Prevalence of disc-related degeneration findings on MR images of the lumbar spine in asymptomatic subjects

Author, year	N	Age mean ± sd (range)	Gender	Bulge	Protrusion	Extrusion	Reduced Signal Intensity	Reduced Disc Height	Annular Tears (HIZ)	Schmorl's nodes
Salo, 1995[173]	49	8 (0-14)	na	-	-	-	22%	-	-	-
Tertti, 1991[14]	39	15	44% M		3%		26%	3%	-	8%
Gibson, 1986[12] or 1987,[266]	20	19 (17-21)	50% M	-	-	-	20%	-	-	-
Paajanen, 1989[16]	34	20 ± 1	100% M	-	-	-	35%	-	-	-
Burns 1996[267]	41	26 (21-31)	100% M	0-10% level	0-32%pe level	-	0-24% level	-	-	7-15% level
Weinreb, 1989[123]	41	30 (19-40)	100% F	44%	-	10%	-	-	-	-
Evans, 1989[268]	59	30	52% M	-	-	-	-	37%	-	-
Schellhas, 1996[18]	17	30 (22-54)	na	-	-	0%	23%	-	6%	-
Weishaupt, 1998[269]	60	35 (20-50)	50% M	20-28%	38-42%	18%	-	-	32-33%	-
Boos, 1995[122]	46	36 (20-50)	74% M	51%	63%	13%	-	-	-	-
Stadnik, 1998[175]	36	42 (17-71)	56% M	81%	33%	-	55%	-	56%	-
Boden, 1990[146]	67	42 (20-80)	45% M	-	59%	24%	-	-	-	-
Boden, 1996[270] (L3S1)	67	42 (20-79)	na	22% discs			54%	-	9% discs	-
Jensen, 1994[151]	98	42 (20-80)	51% M	52%	27%	1%	-	-	14%	19%
Healy, 1996[170]	19	53 (41-69)	100% M	-	58%	21%	-	-	-	-
Jarvik, 2001[152]	148	54 (36-71)	78% M	64%	32%	6%	83%	56%	38%	-
Paajanen, 1997[172]	216	(10-49)	51% M	-	-	-	44%	-	-	-
Parkkola 1993[168]	60	(30-47)	na	15% bpe	-	-	-	-	-	-
Danielson, 2001[271]	43	(20-60)	49% M	-	26%	-	-	-	-	-
Hamanishi, 1994[272]	106	(1-82)	na	-	-	-	-	-	-	9%
Powell 1986[4]	302	(16-80)	100% F	11-13% bpe	-	-	6-79% age	-	-	-

Abbreviations: M = Males, F= Females, na = not available, % disc = % from discs studied, % level = % of subjects at a given intervertebral level, % age = % per age strata, bpe = bulges, protrusions or extrusions, pe = protrusions or extrusions

Table 1.2. Prevalence of disc-related degeneration findings on MR images of the lumbar spine in subjects with symptoms from case-control, population-based, or occupational group studies

Author, year	N (N=20+ included)	Age mean ± sd (range)	Gender	Bulge	Protrusion	Extrusion	Reduced Signal Intensity	Reduced Disc Height	Annular Tears (HIZ)	Schmorl's nodes
Salo, 1995 [173]	32 referred for MRI	8 (0-14)	44% M	-	-	-	13%	-	-	-
Kjaer, 2005[273]	98 with LBP within last month	13	38% M	na	1-2%pe level	-	4-17% level	2-32% level	0-3% level	0-6% level
	34 had seeked care for LBP	13	14% M		3-9%pe level		3-26% level	0-41% level	0-9% level	0-9% level
Tertti, 1991[14]	39 continuous or recurrent LBP	15	44% M	-	21%	-	38%	3%	-	23%
Paajanen, 1989[16]	75 with LBP	20 ± 1	100% M	-	-	-	57%	-	-	-
Boos, 1995[122]	46 with LBP and sciatica scheduled for discectomy	35 (20-50)	74% M	-	54%	35% (7% sequest)	-	-	-	-
Carragee, 2000[43]	42 undergoing discography 64% with previous surgery	36 (22-57)	60% M	-	-	-	-	-	60%	-
Paajanen, 1997[172]	207 with LBP	(10-49)	57% M	-	-	-	59%	-	-	-
Parkkola, 1993[168]	38 moderate LBP	(30-47)	53% M	37%bpe	-	-	-	-	-	-
	10 severe LBP		40% M	90%bpe						
Hamanishi, 1994[272]	400 with MRI for LBP or sciatica	(1-82)	na	-	-	-	-	-	-	19%

Table 1.3. Prevalence of disc-related degeneration findings on MR images of the lumbar spine in population-based or occupational group samples selected without regards to low back pain

Author, year	N, Referral source	Age mean ± sd (range)	Gender	Bulge	Protrusion	Extrusion	Reduced Signal Intensity	Reduced Disc Height	Annular Tears (HIZ)	Schmorl's nodes	Osteophytes
Tertti, 1990[274]	45	12 (8-19)	51% M	-	-	-	9%	-	-	-	-
Kjaer, 2005[273]	439	13 (12-14)	47% M	14%	3%	0%	21%	38%	7%	7%	-
Harcastle, 1992[275]	37	18 (16-18)	na	-	0% pe	-	-	-	-	-	-
Bartolozzi, 1991[276]	30 swimmer, 45 volleyball player	21 (17-26)	63% M 47% M	-	7% 24%	3% 20%	10% 18%	-	-	-	-
Sward, 1991[277]	24 elite gymnasts 16 non athletes	23 (19-29) 26 (23-36)	100%M	50% bpe 19% bpe	-	-	67% 31%	38% 13%	-	71% 44%	-
Kjaer, 2005 [156]	412	40	48% M	28%	23%	1.2% (0.2 extr)	45%	53%	40%	30%	-
Luoma, 1998[255]	53 drivers, 51 carpenters, 60 office workers	(40-45)		0-31% level 12-48% level 5-36% level	-	-	13-76% level 8-68% level 13-62% level	-	-	-	-
Battié, 1995[184]	40 (20 twin pairs discordant for smoking)	46 (35-60)	100% M	-	5-70% pe level	-	10-85% level	13-53% level	-	5-15% level	-
Videman, 2003[278]	230	49 (35-69)	100% M	-	1% disc	2% disc	-	15% disc	15% disc	6% disc	15% disc
Videman, 1995[150]	232	49 (35-69)	100% M	16-34% level	-	0-1% level	42-86% level	5-36% level	-	8-13% level	12-18% level
Videman, 2000[279]	18 rally drivers 14 from population	43 ± 10 55 ± 10	100% M	-	-	55% 50%	-	-	-	67% 79%	6% 57%
Videman, 1995[280]	94 (24 runners, 26 soccer, 19 weight lifters, 25 shooters)	58 (45-68)	100% M	-	-	0.8-4.2%	-	-	-	-	-
Savage, 1997[125]	149 (24 ambulance, 16 hospital porters, 40 car production, 12 draymen, 57 office)	(20-58)	100% M	22%	10%	-	-	-	-	-	-

Abbreviations: M = Males, F= Females, na = not available, % disc = % from discs studied, % level = % of subjects at a given intervertebral level, % age = % per age strata, pe = protrusions, extrusions

REFERENCES - CHAPTER 1

1. Mixter WJ, Barr JS. Rupture of the intervertebral disc with involvement of the spinal canal. *N Engl J Med.* 1934;1934(211):210-215.
2. CSST. *Annexe statistique au rapport annuel d'activité 1997 CSST.* Québec, Bibliothèque Nationale du Québec. 1998: 0-159
3. Spitzer WO, Abenhaim L, Rossignol M et al. Scientific approach to the assessment and management of activity-related spinal disorders. A monograph for clinicians. Report of the Québec Task Force on Spinal Disorders. *Spine.* 1987;12(7 Suppl):S1-59
4. Powell MC, Wilson M, Szypryt P et al. Prevalence of lumbar disc degeneration observed by magnetic resonance in symptomless women. *Lancet.* 1986;2(8520):1366-1367.
5. Deyo RA, Tsui-Wu YJ. Descriptive epidemiology of low-back pain and its related medical care in the United States. *Spine.* 1987;12(3):264-268.
6. Castagnera L, Grenier N, Lavignolle B et al. Study of correlation between intradiscal pressure and magnetic resonance imaging data in evaluation of disc degeneration. Therapeutic issue with percutaneous nucleotomy. *Spine.* 1991;16(3):348-352.
7. Borenstein D. Epidemiology, etiology, diagnostic evaluation, and treatment of low back pain. *Curr Opin Rheumatol.* 1996;8(2):124-129.
8. Frymoyer JW, Cats-Baril WL. An overview of the incidences and costs of low back pain. *Orthop Clin North Am.* 1991;22(2):263-271.
9. Cassidy JD, Carroll LJ, Cote P. The Saskatchewan health and back pain survey. The prevalence of low back pain and related disability in Saskatchewan adults. *Spine.* 1998;23(17):1860-1866.
10. Hart LG, Deyo RA, Cherkin DC. Physician office visits for low back pain. Frequency, clinical evaluation, and treatment patterns from a U.S. national survey. *Spine.* 1995;20(1):11-19.
11. Borenstein DG. Epidemiology, etiology, diagnostic evaluation, and treatment of low back pain. *Curr Opin Rheumatol.* 2001;13(2):128-134.
12. Gibson MJ, Buckley JH, Worthington BS et al. Magnetic resonance imaging of the spine in adolescents. *Acta Radiol Suppl.* 1986;369:717-718.
13. Andersson GB. The Epidemiology of Spinal Disorders. In: Frymoyer JW, editor. *The Adult Spine: Principles and Practice.* Philadelphia, PA: Lippincott-Raven, 1997: 93-141.
14. Tertti MO, Salminen JJ, Paajanen HE et al. Low-back pain and disk degeneration in children: a case-control MR imaging study. *Radiology.* 1991;180(2):503-507.
15. Frank JW, Kerr MS, Brooker AS et al. Disability resulting from occupational low back pain. Part I: What do we know about primary prevention? A review of the scientific evidence on prevention before disability begins. *Spine.* 1996;21(24):2908-2917.
16. Paajanen H, Erkintalo M, Kuusela T et al. Magnetic resonance study of disc degeneration in young low-back pain patients. *Spine.* 1989;14(9):982-985.

17. Deyo RA, Phillips WR. Low back pain. A primary care challenge. *Spine.* 1996;21(24):2826-2832.
18. Schellhas KP, Pollei SR, Gundry CR et al. Lumbar disc high-intensity zone. Correlation of magnetic resonance imaging and discography. *Spine.* 1996;21(1):79-86.
19. Woodwell DA, Cherry DK. National Ambulatory Medical Care Survey: 2002 Summary. *Advance Data From Vital and Health Statistics.* 2004;(346):1-44.
20. Infante-Rivard C, Lortie M. Relapse and short sickness absence for back pain in the six months after return to work. *Occup Environ Med.* 1997;54(5):328-334.
21. Abenhaim L, Suissa S, Rossignol M. Risk of recurrence of occupational back pain over three year follow up. *Br J Ind Med.* 1988;45(12):829-833.
22. Frymoyer JW. Magnitude of the problem. In: Wiesel SW, Weinstein JN, Herkowitz H, Dvorak J, Bell G, editors. *The Lumbar Spine*. Philadelphia, Pensylvania: W.B. Saunders Company, 1996: 8-16.
23. Damkot DK, Pope MH, Lord J et al. The relationship between work history, work environment and low-back pain in men. *Spine.* 1984;9(4):395-399.
24. Merskey H, Bogduk N. *Classification for chronic pain syndromes and definitions of pain terms*. Second ed. Seattle, Washington: IASP Press, 1994,1-222.
25. Cedraschi C, Robert J, Goerg D et al. Is chronic non-specific low back pain chronic? Definitions of a problem and problems of a definition. *Br J Gen Pract.* 1999;49(442):358-362.
26. Abenhaim L, Rossignol M, Gobeille D et al. The prognostic consequences in the making of the initial medical diagnosis of work-related back injuries. *Spine.* 1995;20(7):791-795.
27. Borkan JM, Koes B, Reis S et al. A report from the Second International Forum for Primary Care Research on Low Back Pain. Reexamining priorities. *Spine.* 1998;23(18):1992-1996.
28. Borkan JM, Cherkin DC. An agenda for primary care research on low back pain. *Spine.* 1996;21(24):2880-2884.
29. Parent E. Highlights from the Alberta International Forum VII: Primary Care Research on Low-Back Pain. *University of Alberta Health Science Journal.* 2005;2(1):46-49.
30. Bogduk N. Low Back Pain. In: Bogduk N, editor. *Clinical Anatomy of the Lumbar Spine and Sacrum*. New York, NY: Churchill Livingstone, 1997: 187-213.
31. McKenzie R, May S. Pain and Connective Tissue Properties. In: Writer's Ink, editor. *The Lumbar Spine Mechanical Diagnosis & Therapy*. Waikanae, New Zealand: Spinal Publications New Zealand Ltd, 2003: 45-66.
32. Kuslich SD, Ulstrom CL, Michael CJ. The tissue origin of low back pain and sciatica: a report of pain response to tissue stimulation during operations on the lumbar spine using local anesthesia. *Orthop Clin North Am.* 1991;22(2):181-187.
33. Schwarzer AC, Aprill CN, Derby R et al. The prevalence and clinical features of internal disc disruption in patients with chronic low back pain. *Spine.* 1995;20(17):1878-1883.
34. Simmons JW, Aprill CN, Dwyer AP et al. A reassessment of Holt's data on: "The question of lumbar discography". *Clin Orthop.* 1988;(237):120-124.

35. Walsh TR, Weinstein JN, Spratt KF et al. Lumbar discography in normal subjects. A controlled, prospective study. *J Bone Joint Surg Am.* 1990;72(7):1081-1088.
36. Vanharanta H, Sachs BL, Spivey MA et al. The relationship of pain provocation to lumbar disc deterioration as seen by CT/discography. *Spine.* 1987;12(3):295-298.
37. Moneta GB, Videman T, Kaivanto K et al. Reported pain during lumbar discography as a function of anular ruptures and disc degeneration. A re-analysis of 833 discograms. *Spine.* 1994;19(17):1968-1974.
38. Milette PC, Fontaine S, Lepanto L et al. Differentiating lumbar disc protrusions, disc bulges, and discs with normal contour but abnormal signal intensity. Magnetic resonance imaging with discographic correlations. *Spine.* 1999;24(1):44-53.
39. Donelson R, Aprill C, Medcalf R et al. A prospective study of centralization of lumbar and referred pain. A predictor of symptomatic discs and anular competence. *Spine.* 1997;22(10):1115-1122.
40. Ohnmeiss DD, Vanharanta H, Ekholm J. Degree of disc disruption and lower extremity pain. *Spine.* 1997;22(14):1600-1605.
41. Milette PC, Fontaine S, Lepanto L et al. Radiating pain to the lower extremities caused by lumbar disk rupture without spinal nerve root involvement. *AJNR Am J Neuroradiol.* 1995;16(8):1605-1613.
42. Falconer MA, McGeorge M, Begg AC. Observations on the Cause and Mechanism of Symptom-Production in Sciatica and Low-Back Pain. *J Neurol Neurosurg & Psychiat.* 1947;11:13-26.
43. Carragee EJ, Paragioudakis SJ, Khurana S. 2000 Volvo Award winner in clinical studies: Lumbar high-intensity zone and discography in subjects without low back problems. *Spine.* 2000;25(23):2987-2992.
44. Carragee EJ, Tanner CM, Khurana S et al. The rates of false-positive lumbar discography in select patients without low back symptoms. *Spine.* 2000;25(11):1373-1380.
45. Mooney V. Presidential address. International Society for the Study of the Lumbar Spine. Dallas, 1986. Where is the pain coming from? *Spine.* 1987;12(8):754-759.
46. Mooney V. Where is the lumbar pain coming from? *Ann Med.* 1989;21(5):373-379.
47. Ackerman SJ, Steinberg EP, Bryan RN et al. Persistent low back pain in patients suspected of having herniated nucleus pulposus: radiologic predictors of functional outcome-- implications for treatment selection. *Radiology.* 1997;203(3):815-822.
48. Garfin SR, Rydevik BL, Brown RA. Compressive neuropathy of spinal nerve roots. A mechanical or biological problem? *Spine.* 1991;16(2):162-166.
49. Olmarker K. Mechanical and biochemical injury of spinal nerve roots: an experimental perspective. In: Weinstein JN, Gordon SL, editors. *Low back pain A scientific and clinical overview.* Rosemont, Illinois: American Academy of Orthopaedic Surgeons, 1996: 215-233.
50. Sugawara O, Atsuta Y, Iwahara T et al. The effects of mechanical compression and hypoxia on nerve root and dorsal root ganglia. An analysis of ectopic firing using an in vitro model. *Spine.* 1996;21(18):2089-2094.
51. Garfin SR, Rydevik B, Lind B et al. Spinal nerve root compression. *Spine.* 1995;20(16):1810-1820.

52. Cornefjord M, Olmarker K, Rydevik R et al. Mechanical and biochemical injury of spinal nerve roots: a morphological and neurophysiological study. *Eur Spine J.* 1996;5(3):187-192.
53. Smyth MJ, Wright V. Sciatica and the intervertebral disc. *J Bone Joint Surg Am.* 1958;40(6):1401-1418.
54. McKenzie R, May S. The Intervertebral Disc. In: Writer's Ink, editor. *The Lumbar Spine Mechanical Diagnosis & Therapy*. Waikanae, New Zealand: Spinal Publications New Zealand Ltd, 2003: 67-86.
55. Kang JD, Stefanovic-Racic M, McIntyre LA et al. Toward a biochemical understanding of human intervertebral disc degeneration and herniation. Contributions of nitric oxide, interleukins, prostaglandin E2, and matrix metalloproteinases. *Spine.* 1997;22(10):1065-1073.
56. Takahashi H, Suguro T, Okazima Y et al. Inflammatory cytokines in the herniated disc of the lumbar spine. *Spine.* 1996;21(2):218-224.
57. Gronblad M, Virri J, Tolonen J et al. A controlled immunohistochemical study of inflammatory cells in disc herniation tissue. *Spine.* 1994;19(24):2744-2751.
58. Nygaard OP, Mellgren SI, Osterud B. The inflammatory properties of contained and noncontained lumbar disc herniation. *Spine.* 1997;22(21):2484-2488.
59. Virri J, Gronblad M, Seitsalo S et al. Comparison of the prevalence of inflammatory cells in subtypes of disc herniations and associations with straight leg raising. *Spine.* 2001;26(21):2311-2315.
60. Ahn SH, Cho YW, Ahn MW et al. mRNA expression of cytokines and chemokines in herniated lumbar intervertebral discs. *Spine.* 2002;27(9):911-917.
61. Haro H, Shinomiya K, Komori H et al. Upregulated expression of chemokines in herniated nucleus pulposus resorption. *Spine.* 1996;21(14):1647-1652.
62. Miyamoto H, Saura R, Doita M et al. The role of cyclooxygenase-2 in lumbar disc herniation. *Spine.* 2002;27(22):2477-2483.
63. Burke JG, Watson RW, McCormack D et al. Intervertebral discs which cause low back pain secrete high levels of proinflammatory mediators. *J Bone Joint Surg Br.* 2002;84(2):196-201.
64. Starkweather A, Witek-Janusek L, Mathews HL. Neural-immune interactions: implications for pain management in patients with low-back pain and sciatica. *Biol Res Nurs.* 2005;6(3):196-206.
65. Aoki Y, Rydevik B, Kikuchi S et al. Local application of disc-related cytokines on spinal nerve roots. *Spine.* 2002;27(15):1614-1617.
66. Coppes MH, Marani E, Thomeer RT et al. Innervation of "painful" lumbar discs. *Spine.* 1997;22(20):2342-2349.
67. Harrington JF, Messier AA, Bereiter D et al. Herniated lumbar disc material as a source of free glutamate available to affect pain signals through the dorsal root ganglion. *Spine.* 2000;25(8):929-936.
68. Yoshida M, Nakamura T, Sei A et al. Intervertebral disc cells produce tumor necrosis factor alpha, interleukin-1beta, and monocyte chemoattractant protein-1 immediately after herniation: an experimental study using a new hernia model. *Spine.* 2005;30(1):55-61.

69. Kanayama M, Hashimoto T, Shigenobu K et al. Efficacy of serotonin receptor blocker for symptomatic lumbar disc herniation. *Clin Orthop Relat Res.* 2003;(411):159-165.
70. Komori H, Shinomiya K, Nakai O et al. The natural history of herniated nucleus pulposus with radiculopathy. *Spine.* 1996;21(2):225-229.
71. Haro H, Crawford HC, Fingleton B et al. Matrix metalloproteinase-7-dependent release of tumor necrosis factor- alpha in a model of herniated disc resorption. *J Clin Invest.* 2000;105(2):143-150.
72. Komori H, Okawa A, Haro H et al. Contrast-enhanced magnetic resonance imaging in conservative management of lumbar disc herniation. *Spine.* 1998;23(1):67-73.
73. Fagerlund MK, Thelander U, Friberg S. Size of lumbar disc hernias measured using computed tomography and related to sciatic symptoms. *Acta Radiol.* 1990;31(6):555-558.
74. Masui T, Yukawa Y, Nakamura S et al. Natural History of Patients with Lumbar Disc Herniation Observed by Magnetic Resonance Imaging for Minimum 7 Years. *J Spinal Disord Tech.* 2005;18(2):121-126.
75. McCarron RF, Wimpee MW, Hudkins PG et al. The inflammatory effect of nucleus pulposus. A possible element in the pathogenesis of low-back pain. *Spine.* 1987;12(8):760-764.
76. Bobechko WP, Hirsh C. Auto-Immune response to nucleus pulposus in the rabbit. *J Bone Joint Surg Br.* 1965;47(3):574-580.
77. Gertzbein SD, Tile M, Gross A et al. Autoimmunity in degenerative disc disease of the lumbar spine. *Orthop Clin North Am.* 1975;6(1):67-73.
78. Marshall LL, Trethewie ER, Curtain CC. Chemical radiculitis. A clinical, physiological and immunological study. *Clin Orthop.* 1977;(129):61-67.
79. Saal JS. The role of inflammation in lumbar pain. *Spine.* 1995;20(16):1821-1827.
80. Brisby H, Balague F, Schafer D et al. Glycosphingolipid antibodies in serum in patients with sciatica. *Spine.* 2002;27(4):380-386.
81. Geiss A, Larsson K, Rydevik B et al. Autoimmune Properties of Nucleus Pulposus: an Experimental Study in Pigs. *Abstracts from the 32nd Annual Meeting of the International Society for the Study of the Lumbar Spine.* 1,15. 2005.
82. Olmarker K, Nordborg C, Larsson K et al. Ultrastructural changes in spinal nerve roots induced by autologous nucleus pulposus. *Spine.* 1996;21(4):411-414.
83. Iwabuchi M, Rydevik B, Kikuchi S et al. Effects of anulus fibrosus and experimentally degenerated nucleus pulposus on nerve root conduction velocity: relevance of previous experimental investigations using normal nucleus pulposus. *Spine.* 2001;26(15):1651-1655.
84. Kawaguchi S, Yamashita T, Yokogushi K et al. Immunophenotypic analysis of the inflammatory infiltrates in herniated intervertebral discs. *Spine.* 2001;26(11):1209-1214.
85. Harrington JF, Messier AA, Hoffman L et al. Physiological and behavioral evidence for focal nociception induced by epidural glutamate infusion in rats. *Spine.* 2005;30(6):606-612.
86. Marshall LL, Trethewie ER. Chemical irritation of nerve-root in disc prolapse. *Lancet.* 1973;2(7824):320.

87. Nachemson A. Intradiscal measurements of pH in patients with lumbar rhizopathies. *Acta Orthop Scand.* 1969;40(1):23-42.
88. Freemont AJ, Peacock TE, Goupille P et al. Nerve ingrowth into diseased intervertebral disc in chronic back pain. *Lancet.* 1997;350(9072):178-181.
89. Furusawa N, Baba H, Miyoshi N et al. Herniation of cervical intervertebral disc: immunohistochemical examination and measurement of nitric oxide production. *Spine.* 2001;26(10):1110-1116.
90. Konttinen YT, Gronblad M, Antti-Poika I et al. Neuroimmunohistochemical analysis of peridiscal nociceptive neural elements. *Spine.* 1990;15(5):383-386.
91. Manchikanti L, Boswell MV, Singh V et al. Prevalence of facet joint pain in chronic spinal pain of cervical, thoracic, and lumbar regions. *BMC Musculoskelet Disord.* 2004;5(1):15.
92. Schwarzer AC, Wang SC, Bogduk N et al. Prevalence and clinical features of lumbar zygapophysial joint pain: a study in an Australian population with chronic low back pain. *Ann Rheum Dis.* 1995;54(2):100-106.
93. Manchikanti L, Singh V, Pampati V et al. Evaluation of the Relative Contributions of Various Structures in Chronic Low Back Pain. *Pain Physician.* 2001;4(4):308-316.
94. Schwarzer AC, Aprill CN, Derby R et al. Clinical features of patients with pain stemming from the lumbar zygapophysial joints. Is the lumbar facet syndrome a clinical entity? *Spine.* 1994;19(10):1132-1137.
95. Yong-Hing K, Kirkaldy-Willis WH. The pathophysiology of degenerative disease of the lumbar spine. *Orthop Clin North Am.* 1983;14(3):491-504.
96. Sims JA, Moorman SJ. The role of the iliolumbar ligament in low back pain. *Med Hypotheses.* 1996;46(6):511-515.
97. Strobel ES, Krapf M, Suckfull M et al. Tissue oxygen measurement and 31P magnetic resonance spectroscopy in patients with muscle tension and fibromyalgia. *Rheumatol Int.* 1997;16(5):175-180.
98. Njoo KH, van der DE. The occurrence and inter-rater reliability of myofascial trigger points in the quadratus lumborum and gluteus medius: a prospective study in non-specific low back pain patients and controls in general practice. *Pain.* 1994;58(3):317-323.
99. Jull GA, Richardson CA. Motor control problems in patients with spinal pain: a new direction for therapeutic exercise. *J Manipulative Physiol Ther.* 2000;23(2):115-117.
100. Hides JA, Richardson CA, Jull GA. Multifidus muscle recovery is not automatic after resolution of acute, first-episode low back pain. *Spine.* 1996;21(23):2763-2769.
101. Hides JA, Stokes MJ, Saide M et al. Evidence of lumbar multifidus muscle wasting ipsilateral to symptoms in patients with acute/subacute low back pain. *Spine.* 1994;19(2):165-172.
102. Maigne JY, Aivaliklis A, Pfefer F. Results of sacroiliac joint double block and value of sacroiliac pain provocation tests in 54 patients with low back pain. *Spine.* 1996;21(16):1889-1892.
103. Schwarzer AC, Aprill CN, Bogduk N. The sacroiliac joint in chronic low back pain. *Spine.* 1995;20(1):31-37.

104. Jackson HC, Winkelmann RK, Bickel WH. Nerve endings in the human lumbar spinal column and related structures. *J Bone Joint Surg Am.* 1966;48(7):1272-1281.
105. Wiltse LL, Fonseca AS, Amster J et al. Relationship of the dura, Hofmann's ligaments, Batson's plexus, and a fibrovascular membrane lying on the posterior surface of the vertebral bodies and attaching to the deep layer of the posterior longitudinal ligament. An anatomical, radiologic, and clinical study. *Spine.* 1993;18(8):1030-1043.
106. Cooper RG, Freemont AJ, Hoyland JA et al. Herniated intervertebral disc-associated periradicular fibrosis and vascular abnormalities occur without inflammatory cell infiltration. *Spine.* 1995;20(5):591-598.
107. Castro WH, Assheuer J, Schulitz KP. Haemodynamic changes in lumbar nerve root entrapment due to stenosis and/or herniated disc of the lumbar spinal canal--a magnetic resonance imaging study. *Eur Spine J.* 1995;4(4):220-225.
108. Berlemann U, Gries NC, Moore RJ. The relationship between height, shape and histological changes in early degeneration of the lower lumbar discs. *Eur Spine J.* 1998;7(3):212-217.
109. Schiebler ML, Camerino VJ, Fallon MD et al. In vivo and ex vivo magnetic resonance imaging evaluation of early disc degeneration with histopathologic correlation. *Spine.* 1991;16(6):635-640.
110. Modic MT, Masaryk TJ, Ross JS et al. Imaging of degenerative disk disease. *Radiology.* 1988;168(1):177-186.
111. Milette PC. Classification, diagnostic imaging, and imaging characterization of a lumbar herniated disk. *Radiol Clin North Am.* 2000;38(6):1267-1292.
112. Sachs BL, Vanharanta H, Spivey MA et al. Dallas discogram description. A new classification of CT/discography in low-back disorders. *Spine.* 1987;12(3):287-294.
113. Aprill C, Bogduk N. High-intensity zone: a diagnostic sign of painful lumbar disc on magnetic resonance imaging. *Br J Radiol.* 1992;65(773):361-369.
114. Bender A, Elstner M, Paul R et al. Severe symptomatic aseptic chemical meningitis following myelography: the role of procalcitonin. *Neurology.* 2004;63(7):1311-1313.
115. Ricketson R, Simmons JW, Hauser BO. The prolapsed intervertebral disc. The high-intensity zone with discography correlation. *Spine.* 1996;21(23):2758-2762.
116. Poynton AR, Hinman A, Lutz G et al. Discography-induced acute lumbar disc herniation: a report of five cases. *J Spinal Disord Tech.* 2005;18(2):188-192.
117. Huang TS, Zucherman JF, Hsu KY et al. Gadopentetate dimeglumine as an intradiscal contrast agent. *Spine.* 2002;27(8):839-843.
118. Naish C, Mitchell R, Innes J et al. Ultrasound imaging of the intervertebral disc. *Spine.* 2003;28(2):107-113.
119. McNally DS, Naish C, Halliwell M. Intervertebral disc structure: observation by a novel use of ultrasound imaging. *Ultrasound Med Biol.* 2000;26(5):751-758.
120. Evliyaoglu C, Ayberk G, Sargin H et al. Diagnostic value of bed-side ultrasonographic evaluation by anterior approach in spine disorders? A preliminary report. *Neurocirugia (Astur).* 2003;14(5):392-397.
121. Beattie P. The relationship between symptoms and abnormal magnetic resonance images of lumbar intervertebral disks. *Phys Ther.* 1996;76(6):601-608.

122. Boos N, Rieder R, Schade V et al. 1995 Volvo Award in clinical sciences. The diagnostic accuracy of magnetic resonance imaging, work perception, and psychosocial factors in identifying symptomatic disc herniations. *Spine.* 1995;20(24):2613-2625.
123. Weinreb JC, Wolbarsht LB, Cohen JM et al. Prevalence of lumbosacral intervertebral disk abnormalities on MR images in pregnant and asymptomatic nonpregnant women. *Radiology.* 1989;170(1 Pt 1):125-128.
124. Szypryt EP, Twining P, Mulholland RC et al. The prevalence of disc degeneration associated with neural arch defects of the lumbar spine assessed by magnetic resonance imaging. *Spine.* 1989;14(9):977-981.
125. Savage RA, Whitehouse GH, Roberts N. The relationship between the magnetic resonance imaging appearance of the lumbar spine and low back pain, age and occupation in males. *Eur Spine J.* 1997;6(2):106-114.
126. Kerttula LI, Serlo WS, Tervonen OA et al. Post-traumatic findings of the spine after earlier vertebral fracture in young patients: clinical and MRI study. *Spine.* 2000;25(9):1104-1108.
127. Jarvik JG, Deyo RA. Imaging of lumbar intervertebral disk degeneration and aging, excluding disk herniations. *Radiol Clin North Am.* 2000;38(6):1255-66, vi.
128. Beattie PF, Meyers SP, Stratford P et al. Associations between patient report of symptoms and anatomic impairment visible on lumbar magnetic resonance imaging. *Spine.* 2000;25(7):819-828.
129. Kim KY, Kim YT, Lee CS et al. MRI classification of lumbar herniated intervertebral disc. *Orthopedics.* 1992;15(4):493-497.
130. Modic MT, Pavlicek W, Weinstein MA et al. Magnetic resonance imaging of intervertebral disk disease. Clinical and pulse sequence considerations. *Radiology.* 1984;152(1):103-111.
131. Georgy BA, Hesselink JR. MR imaging of the spine: recent advances in pulse sequences and special techniques. *AJR Am J Roentgenol.* 1994;162(4):923-934.
132. Rankine JJ, Hutchinson CE, Hughes DG. MRI of lumbar spondylosis: a comparison of sagittal T2 weighted and three sequence examinations. *Br J Radiol.* 1997;70(839):1112-1121.
133. Maravilla KR, Lesh P, Weinreb JC et al. Magnetic resonance imaging of the lumbar spine with CT correlation. *AJNR Am J Neuroradiol.* 1985;6(2):237-245.
134. Ebeling U, Reulen HJ. Are there typical localisations of lumbar disc herniations? A prospective study. *Acta Neurochir (Wien).* 1992;117(3-4):143-148.
135. Fardon DF, Milette PC. Nomenclature and classification of lumbar disc pathology. Recommendations of the Combined task Forces of the North American Spine Society, American Society of Spine Radiology, and American Society of Neuroradiology. *Spine.* 2001;26(5):E93-E113.
136. McCall I, O'Brien J. Plain films of the lumbar spine. In: Wiesel SW, Weinstein JN, Herkowitz H, Dvorak J, Bell G, editors. *The lumbar spine.* Philadelphia, PA: Saunders, 1996: 333-352.
137. Vernon-Roberts B, Pirie CJ. Degenerative changes in the intervertebral discs of the lumbar spine and their sequelae. *Rheumatol Rehabil.* 1977;16(1):13-21.

138. Horton WC, Daftari TK. Which disc as visualized by magnetic resonance imaging is actually a source of pain? A correlation between magnetic resonance imaging and discography. *Spine.* 1992;17(6 Suppl):S164-S171.
139. Luoma K, Vehmas T, Riihimaki H et al. Disc height and signal intensity of the nucleus pulposus on magnetic resonance imaging as indicators of lumbar disc degeneration. *Spine.* 2001;26(6):680-686.
140. Tertti M, Paajanen H, Laato M et al. Disc degeneration in magnetic resonance imaging. A comparative biochemical, histologic, and radiologic study in cadaver spines. *Spine.* 1991;16(6):629-634.
141. Sether LA, Yu S, Haughton VM et al. Intervertebral disk: normal age-related changes in MR signal intensity. *Radiology.* 1990;177(2):385-388.
142. Rankine JJ, Gill KP, Hutchinson CE et al. The clinical significance of the high-intensity zone on lumbar spine magnetic resonance imaging. *Spine.* 1999;24(18):1913-1919.
143. Saifuddin A, Braithwaite I, White J et al. The value of lumbar spine magnetic resonance imaging in the demonstration of anular tears. *Spine.* 1998;23(4):453-457.
144. Smith BM, Hurwitz EL, Solsberg D et al. Interobserver reliability of detecting lumbar intervertebral disc high- intensity zone on magnetic resonance imaging and association of high- intensity zone with pain and anular disruption. *Spine.* 1998;23(19):2074-2080.
145. Pfirrmann CW, Metzdorf A, Zanetti M et al. Magnetic resonance classification of lumbar intervertebral disc degeneration. *Spine.* 2001;26(17):1873-1878.
146. Boden SD, Davis DO, Dina TS et al. Abnormal magnetic-resonance scans of the lumbar spine in asymptomatic subjects. A prospective investigation. *J Bone Joint Surg Am.* 1990;72(3):403-408.
147. Buirski G. Magnetic resonance signal patterns of lumbar discs in patients with low back pain. A prospective study with discographic correlation. *Spine.* 1992;17(10):1199-1204.
148. Tibrewal SB, Pearcy MJ. Lumbar intervertebral disc heights in normal subjects and patients with disc herniation. *Spine.* 1985;10(5):452-454.
149. Brinckmann P, Grootenboer H. Change of disc height, radial disc bulge, and intradiscal pressure from discectomy. An in vitro investigation on human lumbar discs. *Spine.* 1991;16(6):641-646.
150. Videman T, Battie MC, Gill K et al. Magnetic resonance imaging findings and their relationships in the thoracic and lumbar spine. Insights into the etiopathogenesis of spinal degeneration. *Spine.* 1995;20(8):928-935.
151. Jensen MC, Brant-Zawadzki MN, Obuchowski N et al. Magnetic resonance imaging of the lumbar spine in people without back pain. *N Engl J Med.* 1994;331(2):69-73.
152. Jarvik JJ, Hollingworth W, Heagerty P et al. The longitudinal assessment of imaging and disability of the back (laidback) study: baseline data. *Spine.* 2001;26(10):1158-1166.
153. Pfirrmann CW, Dora C, Schmid MR et al. MR image-based grading of lumbar nerve root compromise due to disk herniation: reliability study with surgical correlation. *Radiology.* 2004;230(2):583-588.

154. Mulholland RC, McCall I. Magnetic resonance Imaging of the Lumbar Spine. In: Wiesel SW, Weinstein JN, Herkowitz H, Dvorak J, Bell G, editors. *The Lumbar Spine*. Philadelphia: Saunders, 1996: 353-375.
155. Raininko R, Manninen H, Battie MC et al. Observer variability in the assessment of disc degeneration on magnetic resonance images of the lumbar and thoracic spine. *Spine.* 1995;20(9):1029-1035.
156. Kjaer P, Leboeuf-Yde C, Korsholm L et al. Magnetic resonance imaging and low back pain in adults: a diagnostic imaging study of 40-year-old men and women. *Spine.* 2005;30(10):1173-1180.
157. Brant-Zawadzki MN, Jensen MC, Obuchowski N et al. Interobserver and intraobserver variability in interpretation of lumbar disc abnormalities. A comparison of two nomenclatures. *Spine.* 1995;20(11):1257-1263.
158. Jarvik JG, Haynor DR, Koepsell TD et al. Interreader reliability for a new classification of lumbar disk disease. *Acad Radiol.* 1996;3(7):537-544.
159. Karppinen J, Malmivaara A, Tervonen O et al. Severity of symptoms and signs in relation to magnetic resonance imaging findings among sciatic patients. *Spine.* 2001;26(7):E149-E154.
160. Borenstein DG, O'Mara JW, Jr., Boden SD et al. The value of magnetic resonance imaging of the lumbar spine to predict low-back pain in asymptomatic subjects : a seven-year follow-up study. *J Bone Joint Surg Am.* 2001;83-A(9):1306-1311.
161. Milette PC. The proper terminology for reporting lumbar intervertebral disk disorders. *AJNR Am J Neuroradiol.* 1997;18(10):1859-1866.
162. Porter RW. Does hard work prevent disc protrusion? *Clin Biomech.* 1987;2(4):196-198.
163. Gunzburg R, Parkinson R, Moore R et al. A cadaveric study comparing discography, magnetic resonance imaging, histology, and mechanical behavior of the human lumbar disc. *Spine.* 1992;17(4):417-426.
164. Porter RW, Adams MA, Hutton WC. Physical activity and the strength of the lumbar spine. *Spine.* 1989;14(2):201-203.
165. Yu SW, Haughton VM, Sether LA et al. Anulus fibrosus in bulging intervertebral disks. *Radiology.* 1988;169(3):761-763.
166. Wood KB, Garvey TA, Gundry C et al. Magnetic resonance imaging of the thoracic spine. Evaluation of asymptomatic individuals. *J Bone Joint Surg Am.* 1995;77(11):1631-1638.
167. Boos N, Semmer N, Elfering A et al. Natural history of individuals with asymptomatic disc abnormalities in magnetic resonance imaging: predictors of low back pain-related medical consultation and work incapacity. *Spine.* 2000;25(12):1484-1492.
168. Parkkola R, Rytokoski U, Kormano M. Magnetic resonance imaging of the discs and trunk muscles in patients with chronic low back pain and healthy control subjects. *Spine.* 1993;18(7):830-836.
169. Salminen JJ, Erkintalo M, Laine M et al. Low back pain in the young. A prospective three-year follow-up study of subjects with and without low back pain. *Spine.* 1995;20(19):2101-2107.

170. Healy JF, Healy BB, Wong WH et al. Cervical and lumbar MRI in asymptomatic older male lifelong athletes: frequency of degenerative findings. *J Comput Assist Tomogr*. 1996;20(1):107-112.
171. Battie MC, Videman T, Parent E. Lumbar Disc Degeneration: Epidemiology and Genetic Influences. *Spine*. 2004;29(23):2679-2690.
172. Paajanen H, Erkintalo M, Parkkola R et al. Age-dependent correlation of low-back pain and lumbar disc regeneration. *Arch Orthop Trauma Surg*. 1997;116(1-2):106-107.
173. Salo S, Paajanen H, Alanen A. Disc degeneration of pediatric patients in lumbar MRI. *Pediatr Radiol*. 1995;25(3):186-189.
174. Parent E, Battie MC, Videman T et al. Prevalence of Lumbar Disc Findings on MRI in Persons with Various LBP Histories. *Abstract of the 30th Annual ISSLS meeting*. 224. 5-13-2003.
175. Stadnik TW, Lee RR, Coen HL et al. Annular tears and disk herniation: prevalence and contrast enhancement on MR images in the absence of low back pain or sciatica. *Radiology*. 1998;206(1):49-55.
176. Elfering A, Semmer N, Birkhofer D et al. Young Investigator Award 2001 winner: risk factors for lumbar disc degeneration: a 5-year prospective MRI study in asymptomatic individuals. *Spine*. 2002;27(2):125-134.
177. Gruber HE, Hanley EN, Jr. Recent advances in disc cell biology. *Spine*. 2003;28(2):186-193.
178. Yoon ST. The potential of gene therapy for the treatment of disc degeneration. *Orthop Clin North Am*. 2004;35(1):95-100.
179. Streiner DL, Norman GR. *Health measurement scales: A practical guide to their development and use*. Second ed. New York, New York: Oxford Medical Publications, 1995,1-231.
180. Scientific Advisory Committee of the Medical Outcome Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res*. 2002;11(3):193-205.
181. Glass GV, Hopkins KD. *Statistical methods in education and psychology*. 3 ed. Needham Heights, MA, USA: Allyn and Bacon, 1995,1-674.
182. Battie MC, Videman T, Gibbons LE et al. 1995 Volvo Award in clinical sciences. Determinants of lumbar disc degeneration. A study relating lifetime exposures and magnetic resonance imaging findings in identical twins. *Spine*. 1995;20(24):2601-2612.
183. van Rijn JC, Klemetso N, Reitsma JB et al. Observer variation in MRI evaluation of patients suspected of lumbar disk herniation. *AJR Am J Roentgenol*. 2005;184(1):299-303.
184. Battie MC, Haynor DR, Fisher LD et al. Similarities in degenerative findings on magnetic resonance images of the lumbar spines of identical twins. *J Bone Joint Surg Am*. 1995;77(11):1662-1670.
185. Hollingworth W, Dixon AK, Todd CJ et al. Self reported health status and magnetic resonance imaging findings in patients with low back pain. *Eur Spine J*. 1998;7(5):369-375.

186. Cicchetti DV. Sample size requirements for increasing the precision of reliability estimates: problems and proposed solutions. *Journal of Clinical and Experimental Neuropsychology*. 1999;21(4):567-570.
187. Forristall RM, Marsh HO, Pay NT. Magnetic resonance imaging and contrast CT of the lumbar spine. Comparison of diagnostic methods and correlation with surgical findings. *Spine*. 1988;13(9):1049-1054.
188. Greenspan A, Amparo EG, Gorczyca DP et al. Is there a role for diskography in the era of magnetic resonance imaging? Prospective correlation and quantitative analysis of computed tomography-diskography, magnetic resonance imaging, and surgical findings. *J Spinal Disord*. 1992;5(1):26-31.
189. Birney TJ, White JJ, Jr., Berens D et al. Comparison of MRI and discography in the diagnosis of lumbar degenerative disc disease. *J Spinal Disord*. 1992;5(4):417-423.
190. Janssen ME, Bertrand SL, Joe C et al. Lumbar herniated disk disease: comparison of MRI, myelography, and post-myelographic CT scan with surgical findings. *Orthopedics*. 1994;17(2):121-127.
191. Platenberg RC, Hubbard GB, Ehler WJ et al. Spontaneous disc degeneration in the baboon model: magnetic resonance imaging and histopathologic correlation. *J Med Primatol*. 2001;30(5):268-272.
192. Wittenberg RH, Lutke A, Longwitz D et al. The correlation between magnetic resonance imaging and the operative and clinical findings after lumbar microdiscectomy. *Int Orthop*. 1998;22(4):241-244.
193. Schulz KS, Walker M, Moon M et al. Correlation of clinical, radiographic, and surgical localization of intervertebral disc extrusion in small-breed dogs: a prospective study of 50 cases. *Vet Surg*. 1998;27(2):105-111.
194. Silverman CS, Lenchik L, Shimkin PM et al. The value of MR in differentiating subligamentous from supraligamentous lumbar disk herniations. *AJNR Am J Neuroradiol*. 1995;16(3):571-579.
195. Hashimoto K, Akahori O, Kitano K et al. Magnetic resonance imaging of lumbar disc herniation. Comparison with myelography. *Spine*. 1990;15(11):1166-1169.
196. Grenier N, Greselle JF, Vital JM et al. Normal and disrupted lumbar longitudinal ligaments: correlative MR and anatomic study. *Radiology*. 1989;171(1):197-205.
197. Carragee EJ, Kim DH. A prospective analysis of magnetic resonance imaging findings in patients with sciatica and lumbar disc herniation. Correlation of outcomes with disc fragment and canal morphology. *Spine*. 1997;22(14):1650-1660.
198. Weishaupt D, Schmid MR, Zanetti M et al. Positional MR imaging of the lumbar spine: does it demonstrate nerve root compromise not visible at conventional MR imaging? *Radiology*. 2000;215(1):247-253.
199. Zamani AA, Moriarty T, Hsu L et al. Functional MRI of the lumbar spine in erect position in a superconducting open-configuration MR system: preliminary results. *J Magn Reson Imaging*. 1998;8(6):1329-1333.
200. Splendiani A, Puglielli E, De Amicis R et al. Spontaneous resolution of lumbar disk herniation: predictive signs for prognostic evaluation. *Neuroradiology*. 2004;46(11):916-922.

201. Ahn SH, Park HW, Byun WM et al. Comparison of clinical outcomes and natural morphologic changes between sequestered and large central extruded disc herniations. *Yonsei Med J.* 2002;43(3):283-290.
202. Delauche-Cavallier MC, Budet C, Laredo JD et al. Lumbar disc herniation. Computed tomography scan changes after conservative treatment of nerve root compression. *Spine.* 1992;17(8):927-933.
203. Saal JA, Saal JS, Herzog RJ. The natural history of lumbar intervertebral disc extrusions treated nonoperatively. *Spine.* 1990;15(7):683-686.
204. Ellenberg MR, Ross ML, Honet JC et al. Prospective evaluation of the course of disc herniations in patients with proven radiculopathy. *Arch Phys Med Rehabil.* 1993;74(1):3-8.
205. Gallucci M, Bozzao A, Orlandi B et al. Does postcontrast MR enhancement in lumbar disk herniation have prognostic value? *J Comput Assist Tomogr.* 1995;19(1):34-38.
206. Andersson GB, Buckwalter JA, Deyo RA et al. *Low Back Pain A Scientific and Clinical Overview.* First ed. Rosemont, Illinois: American Academy of Orthopedic Surgeons, 1996,1-766.
207. Teplick JG, Haskin ME. Spontaneous regression of herniated nucleus pulposus.. *AJR Am J Roentgenol.* 1985;145(2):371-375.
208. Kawaji Y, Uchiyama S, Yagi E. Three-dimensional evaluation of lumbar disc hernia and prediction of absorption by enhanced MRI. *J Orthop Sci.* 2001;6(6):498-502.
209. Matsubara Y, Kato F, Mimatsu K et al. Serial changes on MRI in lumbar disc herniations treated conservatively. *Neuroradiology.* 1995;37(5):378-383.
210. Modic MT, Ross JS, Obuchowski NA et al. Contrast-enhanced MR imaging in acute lumbar radiculopathy: a pilot study of the natural history. *Radiology.* 1995;195(2):429-435.
211. Salminen JJ, Erkintalo MO, Pentti J et al. Recurrent low back pain and early disc degeneration in the young. *Spine.* 1999;24(13):1316-1321.
212. Benneker LM, Heini PF, Anderson SE et al. Correlation of radiographic and MRI parameters to morphological and biochemical assessment of intervertebral disc degeneration. *Eur Spine J.* 2005;14(1):27-35.
213. Dai LY. Disc degeneration in patients with lumbar spondylolysis. *J Spinal Disord.* 2000;13(6):478-486.
214. Videman T, Nummi P, Battie MC et al. Digital assessment of MRI for lumbar disc desiccation. A comparison of digital versus subjective assessments and digital intensity profiles versus discogram and macroanatomic findings. *Spine.* 1994;19(2):192-198.
215. Brault JS, Driscoll DM, Laakso LL et al. Quantification of lumbar intradiscal deformation during flexion and extension, by mathematical analysis of magnetic resonance imaging pixel intensity profiles. *Spine.* 1997;22(18):2066-2072.
216. Luoma EK, Raininko R, Nummi PJ et al. Suitability of cerebrospinal fluid as a signal-intensity reference on MRI: evaluation of signal-intensity variations in the lumbosacral dural sac. *Neuroradiology.* 1997;39(10):728-732.

217. Luoma K, Raininko R, Nummi P et al. Is the signal intensity of cerebrospinal fluid constant? Intensity measurements with high and low field magnetic resonance imagers. *Magn Reson Imaging.* 1993;11(4):549-555.
218. Videman T, Leppavuori J, Kaprio J et al. Intragenic polymorphisms of the vitamin D receptor gene associated with intervertebral disc degeneration. *Spine.* 1998;23(23):2477-2485.
219. Amonoo-Kuofi HS. Morphometric changes in the heights and anteroposterior diameters of the lumbar intervertebral discs with age. *J Anat.* 1991;175:159-168.
220. Davies KM, Recker RR, Heaney RP. Normal vertebral dimensions and normal variation in serial measurements of vertebrae. *J Bone Miner Res.* 1989;4(3):341-349.
221. Chen MW, Yang SW, Lee MC. Changes in lumbar disc angles of Chinese subjects from upright to flexion. *Spine.* 1994;19(13):1490-1494.
222. Frobin W, Brinckmann P, Kramer M et al. Height of lumbar discs measured from radiographs compared with degeneration and height classified from MR images. *Eur Radiol.* 2001;11(2):263-269.
223. Murata M, Morio Y, Kuranobu K. Lumbar disc degeneration and segmental instability: a comparison of magnetic resonance images and plain radiographs of patients with low back pain. *Arch Orthop Trauma Surg.* 1994;113(6):297-301.
224. Gilad I, Nissan M. A study of vertebra and disc geometric relations of the human cervical and lumbar spine. *Spine.* 1986;11(2):154-157.
225. Nissan M, Gilad I. Dimensions of human lumbar vertebrae in the sagittal plane. *J Biomech.* 1986;19(9):753-758.
226. Dabbs VM, Dabbs LG. Correlation between disc height narrowing and low-back pain. *Spine.* 1990;15(12):1366-1369.
227. Chen YL. Geometric measurements of the lumbar spine in Chinese men during trunk flexion. *Spine.* 1999;24(7):666-669.
228. Gilad I, Nissan M. Sagittal evaluation of elemental geometrical dimensions of human vertebrae. *J Anat.* 1985;143:115-120.
229. Shao Z, Rompe G, Schiltenwolf M. Radiographic changes in the lumbar intervertebral discs and lumbar vertebrae with age. *Spine.* 2002;27(3):263-268.
230. Boos N, Wallin A, Aebi M et al. A new magnetic resonance imaging analysis method for the measurement of disc height variations. *Spine.* 1996;21(5):563-570.
231. Jaovisidha S, Techatipakorn S, Apiyasawat P et al. Degenerative disk disease at lumbosacral junction: plain film findings and related MRI abnormalities. *J Med Assoc Thai.* 2000;83(8):865-871.
232. Kimura S, Steinbach GC, Watenpaugh DE et al. Lumbar spine disc height and curvature responses to an axial load generated by a compression device compatible with magnetic resonance imaging. *Spine.* 2001;26(23):2596-2600.
233. Kourtis D, Magnusson ML, Smith F et al. Spine height and disc height changes as the effect of hyperextension using stadiometry and MRI. *Iowa Orthop J.* 2004;24:65-71.
234. Edmondston SJ, Song S, Bricknell RV et al. MRI evaluation of lumbar spine flexion and extension in asymptomatic individuals. *Man Ther.* 2000;5(3):158-164.

235. Lee J, Hida K, Seki T et al. An interspinous process distractor (X STOP) for lumbar spinal stenosis in elderly patients: preliminary experiences in 10 consecutive cases. *J Spinal Disord Tech.* 2004;17(1):72-77.

236. Chung SS, Lee CS, Kim SH et al. Effect of low back posture on the morphology of the spinal canal. *Skeletal Radiol.* 2000;29(4):217-223.

237. Botsford DJ, Esses SI, Ogilvie-Harris DJ. In vivo diurnal variation in intervertebral disc volume and morphology. *Spine.* 1994;19(8):935-940.

238. Malko JA, Hutton WC, Fajman WA. An In Vivo MRI Study of the Changes in Volume (and Fluid Content) of the Lumbar Intervertebral Disc After Overnight Bed Rest and During an 8- Hour Walking Protocol. *J Spinal Disord Tech.* 2002;15(2):157-163.

239. Malko JA, Hutton WC, Fajman WA. An in vivo magnetic resonance imaging study of changes in the volume (and fluid content) of the lumbar intervertebral discs during a simulated diurnal load cycle. *Spine.* 1999;24(10):1015-1022.

240. Roberts N, Hogg D, Whitehouse GH et al. Quantitative analysis of diurnal variation in volume and water content of lumbar intervertebral discs. *Clin Anat.* 1998;11(1):1-8.

241. Bozzao A, Gallucci M, Masciocchi C et al. Lumbar disk herniation: MR imaging assessment of natural history in patients treated without surgery. *Radiology.* 1992;185(1):135-141.

242. Holodny AI, Kisza PS, Contractor S et al. Does a herniated nucleus pulposus contribute significantly to a decrease in height of the intervertebral disc? Quantitative volumetric MRI. *Neuroradiology.* 2000;42(6):451-454.

243. Kato F, Mimatsu K, Kawakami N et al. Serial changes observed by magnetic resonance imaging in the intervertebral disc after chemonucleolysis. A consideration of the mechanism of chemonucleolysis. *Spine.* 1992;17(8):934-939.

244. Wood KB, Blair JM, Aepple DM et al. The natural history of asymptomatic thoracic disc herniations. *Spine.* 1997;22(5):525-529.

245. Buttermann GR. Lumbar disc herniation regression after successful epidural steroid injection. *J Spinal Disord Tech.* 2002;15(6):469-476.

246. Tonami H, Yokota H, Nakagawa T et al. Percutaneous laser discectomy: MR findings within the first 24 hours after treatment and their relationship to clinical outcome. *Clin Radiol.* 1997;52(12):938-944.

247. Pneumaticos SG, Hipp JA, Esses SI. Sensitivity and specificity of dural sac and herniated disc dimensions in patients with low back-related leg pain. *J Magn Reson Imaging.* 2000;12(3):439-443.

248. Thelander U, Fagerlund M, Friberg S et al. Describing the size of lumbar disc herniations using computed tomography. A comparison of different size index calculations and their relation to sciatica. *Spine.* 1994;19(17):1979-1984.

249. Ahn SH, Ahn MW, Byun WM. Effect of the transligamentous extension of lumbar disc herniations on their regression and the clinical outcome of sciatica. *Spine.* 2000;25(4):475-480.

250. Henmi T, Sairyo K, Nakano S et al. Natural history of extruded lumbar intervertebral disc herniation. *J Med Invest.* 2002;49(1-2):40-43.

251. Nakano M, Matsui H, Ishihara H et al. Serial changes of herniated intervertebral discs after posterior lumbar discectomy: the relation between magnetic resonance

imaging of the postoperative intervertebral discs and clinical outcome. *J Spinal Disord.* 2001;14(4):293-300.

252. Thelander U, Fagerlund M, Friberg S et al. Straight leg raising test versus radiologic size, shape, and position of lumbar disc hernias. *Spine.* 1992;17(4):395-399.
253. Fries JW, Abodeely DA, Vijungco JG et al. Computed tomography of herniated and extruded nucleus pulposus. *J Comput Assist Tomogr.* 1982;6(5):874-887.
254. Vincent JM, Baldwin JE, Sims C et al. Vertebral 'corner' defect associated with lumbar disk herniation shown by magnetic resonance imaging. *Spine.* 1993;18(1):109-113.
255. Luoma K, Riihimaki H, Raininko R et al. Lumbar disc degeneration in relation to occupation. *Scand J Work Environ Health.* 1998;24(5):358-366.
256. Fredericson M, Lee SU, Welsh J et al. Changes in posterior disc bulging and intervertebral foraminal size associated with flexion-extension movement: a comparison between L4-5 and L5-S1 levels in normal subjects. *Spine J.* 2001;1(1):10-17.
257. Cooley JR, Danielson CD, Schultz GD et al. Posterior disk displacement: morphologic assessment and measurement reliability-lumbar spine. *J Manipulative Physiol Ther.* 2001;24(5):317-326.
258. Inufusa A, An HS, Lim TH et al. Anatomic changes of the spinal canal and intervertebral foramen associated with flexion-extension movement. *Spine.* 1996;21(21):2412-2420.
259. Fujiwara A, An HS, Lim TH et al. Morphologic changes in the lumbar intervertebral foramen due to flexion-extension, lateral bending, and axial rotation: an in vitro anatomic and biomechanical study. *Spine.* 2001;26(8):876-882.
260. Sahlstrand T, Lonntoft M. A prospective study of preoperative and postoperative sequential magnetic resonance imaging and early clinical outcome in automated percutaneous lumbar discectomy. *J Spinal Disord.* 1999;12(5):368-374.
261. Luoma K, Vehmas T, Raininko R et al. Lumbosacral transitional vertebra: relation to disc degeneration and low back pain. *Spine.* 2004;29(2):200-205.
262. Luoma K, Riihimaki H, Luukkonen R et al. Low back pain in relation to lumbar disc degeneration. *Spine.* 2000;25(4):487-492.
263. Silcox DH, III, Horton WC, Silverstein AM. MRI of lumbar intervertebral discs. Diurnal variations in signal intensities. *Spine.* 1995;20(7):807-811.
264. Videman T, Nurminen M, Troup JD. 1990 Volvo Award in clinical sciences. Lumbar spinal pathology in cadaveric material in relation to history of back pain, occupation, and physical loading. *Spine.* 1990;15(8):728-740.
265. Harada A, Okuizumi H, Miyagi N et al. Correlation between bone mineral density and intervertebral disc degeneration. *Spine.* 1998;23(8):857-861.
266. Gibson MJ, Szypryt EP, Buckley JH et al. Magnetic resonance imaging of adolescent disc herniation. *J Bone Joint Surg Br.* 1987;69(5):699-703.
267. Burns JW, Loecker TH, Fischer JR, Jr. et al. Prevalence and significance of spinal disc abnormalities in an asymptomatic acceleration subject panel. *Aviat Space Environ Med.* 1996;67(9):849-853.
268. Evans W, Jobe W, Seibert C. A cross-sectional prevalence study of lumbar disc degeneration in a working population. *Spine.* 1989;14(1):60-64.

269. Weishaupt D, Zanetti M, Hodler J et al. MR imaging of the lumbar spine: prevalence of intervertebral disk extrusion and sequestration, nerve root compression, end plate abnormalities, and osteoarthritis of the facet joints in asymptomatic volunteers. *Radiology.* 1998;209(3):661-666.

270. Boden SD, Riew KD, Yamaguchi K et al. Orientation of the lumbar facet joints: association with degenerative disc disease. *J Bone Joint Surg Am.* 1996;78(3):403-411.

271. Danielson B, Willen J. Axially loaded magnetic resonance image of the lumbar spine in asymptomatic individuals. *Spine.* 2001;26(23):2601-2606.

272. Hamanishi C, Kawabata T, Yosii T et al. Schmorl's nodes on magnetic resonance imaging. Their incidence and clinical relevance. *Spine.* 1994;19(4):450-453.

273. Kjaer P, Leboeuf-Yde C, Sorensen JS et al. An epidemiologic study of MRI and low back pain in 13-year-old children. *Spine.* 2005;30(7):798-806.

274. Tertti M, Paajanen H, Kujala UM et al. Disc degeneration in young gymnasts. A magnetic resonance imaging study. *Am J Sports Med.* 1990;18(2):206-208.

275. Hardcastle P, Annear P, Foster DH et al. Spinal abnormalities in young fast bowlers. *J Bone Joint Surg Br.* 1992;74(3):421-425.

276. Bartolozzi C, Caramella D, Zampa V et al. Incidence of intervertebral disk abnormalities in volley-ball players. MR features. *La radiologia Medica.* 1991;1991(82):757-760.

277. Sward L, Hellstrom M, Jacobsson B et al. Disc degeneration and associated abnormalities of the spine in elite gymnasts. A magnetic resonance imaging study. *Spine.* 1991;16(4):437-443.

278. Videman T, Battie MC, Gibbons LE et al. Associations between back pain history and lumbar MRI findings. *Spine.* 2003;28(6):582-588.

279. Videman T, Simonen R, Usenius J et al. The long-term effects of rally driving on spinal pathology. *Clin Biomech (Bristol, Avon).* 2000;15(2):83-86.

280. Videman T, Sarna S, Battie MC et al. The long-term effects of physical loading and exercise lifestyles on back-related symptoms, disability, and spinal pathology among men. *Spine.* 1995;20(6):699-709.

CHAPTER 2: QUANTITATIVE MRI MEASURES OF LUMBAR DISC CONTOUR ABNORMALITY: DEVELOPMENT AND RELIABILITY

INTRODUCTION

Intervertebral discs have been primary suspects in low back pain (LBP) since Mixter and Barr first reported that disc contour abnormalities were associated with sciatic symptoms.[1] The numerous interventions and diagnostic procedures focussing on these disc findings attest to this association. Typically, radiologist or orthopedists report their visual judgment of the disc's degenerative findings such as disc contour abnormality from magnetic resonance (MR) images using ordinal scale descriptors.

However, the reliability of qualitative grading is suboptimal and varies with rater training or speciality,[2] the scale used,[3-9] the disc level,[2-4,10,11] and the severity of degeneration.[2,4,8,10] In research, low reliability of qualitative grading can mask correlations.[12] This factor could help to explain why controversy still exists regarding the determinants and consequences of disc contour abnormality. A more sensitive and reliable quantitative measure of "disc contour abnormality," used here for bulging or herniation, could help to clarify the strength of the associations, or lack thereof, between contour abnormalities and their suspected determinants and consequences. For example, a significant association between a measure of disc signal intensity and two polymorphisms of the vitamin D receptor gene was revealed when a quantitative measure was used but was undetected by using a less reliable qualitative rating scale.[10,13,14] Quantitative measures could also be ideal for follow-up and correlation studies because these measures have more measurement levels than qualitative rating scales, thereby increasing the likelihood of detecting change.[15]

Reports using quantitative measures of disc contour abnormality are rare. Botsford et al. measured the disc diameter to indicate contour abnormality and found the effect of diurnal variation to be minimal.[16] Any changes, if present, were likely too small relative to the overall disc diameter to be detected. The volume of disc contour abnormalities has been calculated, but the reliability of volume measurments has not been studied adequately.[17,18] The ratio of contour abnormality dimensions to the spinal canal's area on axial images has been assessed, and adequate reports of reliability are sparse, yet promising.[17,19-22] Inter-rater error was estimated to be as low as 2.8% for an average of three measurements.[20,21] However, disc material reportedly migrates caudally or cephally in 65% to 79% of cases;[23-25] thus, axial measures obtained at mid-disc level do not capture the size of migrated disc material.

In the sagittal plane, distance measures of disc contour abnormality[26-28] and ratios of the contour abnormality's dimensions to the spinal canal's area[17,19,29,30] have been used. However, reports on reliability are typically inadequate and descriptions are usually insufficient to allow replication.[18,19,29-32] The reliability estimates reported for such measures vary widely (ICC: 0.40 to 0.92),[28,33] and quantitative measures are applied without reporting their measurement properties in most studies. Sagittal quantitative measures of disc contour abnormalities with high reliability are needed to assist research on their determinants and consequences.

The presence of osteophytes complicates the measure of disc contour abnormality because with osteophytes, identifying reference points from which to obtain a measure is difficult.[2,28,34] To address this issue, Harada et al. suggested a standardized procedure to identify a position from which to measure the area of disc contour abnormality on sagittal images by using the mid-point of the posterior vertebral wall.[34] However, this procedure does not consider that, on mid-sagittal MR images, identifying the position of the posterior vertebral wall is difficult because blood vessels are seen entering the vertebral body at its midpoint.

This study's primary objective was to determine the reliability of the proposed quantitative measures of lumbar disc contour abnormality. The secondary objectives were to (1) compare the reliability of quantitative measures to that of qualitative measures, (2) compare the reliability of distance and area measures of disc contour abnormality, and (3) determine the size of error due to image resolution, intra-, and inter-evaluator differences.

It was hypothesized that the reliability of measures not relying on defining disc contour abnormality from the "corners" of the vertebral bodies where osteophytes commonly form would be higher than that of measures relying on precise segmentation of these "corners" because of the difficulty in defining the vertebra's "corners" on MR images. The reliability of the quantitative measures was expected to be greater than that of the qualitative ratings. The measurement error for high-resolution images was expected to be smaller than that for lower resolution images.

MATERIALS AND METHODS

Subjects

Thirty-one male volunteers with a mean age of 42.2 years (range: 24-74 years) with or without LBP symptoms were recruited by using bulletin board notices. Males were selected because the measures were developed for a large epidemiological study of males. However, reliability results should apply to both genders. Study protocols were approved by the Ethical Committees of the principal investigators' institutions. All subjects received written information about the study and provided written consent before participation.

MR Imaging

Subjects were imaged while lying supine with a pillow supporting their knees. A highly shielded 1.5T Siemens Sonata scanner (Siemens AG, Germany) was used. Turbo spin echo midsagittal lumbar images (T12-S1) were acquired by using the lumbar spine array coils full matrix phase encoding at 256 x 256 and 512 x 512 resolutions (pixel size of 1 mm and 0.5 mm respectively) during one session: proton density (PD) [Time Repetition (TR)/ Time Echo (TE) = 2600/27], T2-[2600/92], T1-[650/26] weighted (NEX = 2 excitations). The total image acquisition time was 1:23:14 hours. Before obtaining the images of interest, subjects spent approximately 30 minutes in the supine position to allow for most of the effect of compression due to weight-bearing on the disc dimensions to normalize.[35] The phase encoding direction was anterior to posterior with an echo train length of 5. The slice thickness was 4 mm with a 0.4 mm interslice gap. The field of view was 260 mm.

Quantitative Assessments

Custom-designed spine image analysis software (SpEx© v2.42 for a PC using Windows operating software) was used to obtain quantitative measures of disc contour abnormality from the midsagittal image only. Images were analysed at a zoom factor of five in their original resolution without application of a smoothing algorithm. The contour of the lumbar discs, vertebrae, and spinal canal was segmented manually on the midsagittal PD image. Segmentation guidelines were developed after examining outliers from a pilot reliability study using a different sample (Figure 2.1) (Appendix C).

One evaluator selected, as midsagittal, the image where the spinous processes of L4 and L5 vertebra were the most clearly revealed. The evaluator segmented the vertical contour of the vertebral wall from the longitudinal ligament. To segment the contour of the disc annulus, at the endplate level, the evaluator moved through the longitudinal ligament to follow the outer contour of the ligament/intervertebral disc. The contrast did not allow the evaluator to distinguish between the annulus and the ligament; thus, the ligament was included in the disc area. To segment the disc from the vertebra, the evaluator followed horizontally the boundary between the vertebral bone and the endplates. Cartilaginous endplates were included in the disc since the contrast did not allow for distinguishing both tissues. The areas created by the intersection of those segmentation lines formed the regions of interest corresponding to the disc and vertebra from which the measures were later derived (Figure 2.1). Segmented regions of interest were copied onto the corresponding midsagittal T2- (disc) and T1-weighted (vertebra) images, manually aligned and adjusted, if necessary, taking advantage of different contrasts.

The distances and areas of disc contour abnormality were calculated using the segmented outlines from two types of reference points (Figure 2.1). A distance and an area of disc contour abnormality was measured by using the "real" vertebral corners (Figure 2.1: R and R') as seen on the MR image at the intersection of the vertebral wall and the endplate contours. The second type of distance and area of disc contour abnormality measures used the vertebrae's "theoretical" corners (Figure 2.1: T and T') as a reference. Theoretical corners are defined by the intersections of a line connecting the diameter of the vertebra above the disc at 20% of the vertebra height to the corresponding diameter below the disc, with the lines representing the disc-vertebra boundaries (Figure 2.1). Measures using theoretical corners were developed to address the difficulty of precisely identifying the position of the real corners on MR images with low contrast and to measure disc contour abnormality when combined with the presence of osteophytes.

Two evaluators manually segmented the images. Approximately 15-20 minutes per image were required. Both evaluators had prior experience segmenting over 200 spine MR images. The first evaluator completed segmentation for the 31 subjects on three occasions at both image resolutions while the second evaluator completed segmentation on one occasion at both resolutions. The participants' images were manually segmented in random order. The evaluators were blind to each other's work, to the participants' medical histories and identities, and to the qualitative raters' assessments. The evaluators completed repeated manual segmentations for a given image approximately a month apart to prevent the inflation of reliability estimates by recall.

Qualitative Assessments

To compare quantitative and qualitative measurements, the reliability of clinicians' qualitative ratings of bulging and herniation was estimated by using the high-resolution sagittal images, typical of the images routinely used in current clinical practice. An orthopedic spine surgeon and a musculoskeletal radiologist rated bulging and herniation at each level from L1 to S1 by using all sagittal images on two separate occasions at least a month apart. These raters had over 10 years of experience reading lumbar MR images. They were blind to the subjects' identities and medical histories as well as to each other's ratings. The images were presented randomly. The qualitative raters had access to all images from the PD, T1-, and T2-weighted sequences in DICOM format by using their preferred image viewing software and could adjust the same display parameters to optimize visualization as quantitative evaluators. Bulging was rated qualitatively as follows: 0 = none, normal contour of the disc; 1 = slight (1.5 ± 1 mm), 2 = moderate (3.5 ± 1 mm), and 3 = severe (≥ 4.5 mm bulge). Herniation was rated as 0 = none; 1 = slight (0-5 mm); 2 = moderate (6-10 mm), and 3 = large (> 1 cm).[36]

Analyses

Reliability was assessed using Generalizability Theory because it provides the following advantages over other assessment methods: (1) Generalizability coefficients (GcoefS) are preferred to correlation coefficients because they measure agreement and not simply associations and thus, do not overestimate the reliability when a systematic difference is present[37], (2) G-theory has the flexibility to quantify the effect of more than one source of error in a single analysis which is not possible using classical test theory[38p.106-7] statistics such as intraclass correlation coefficients[39], coefficients of variations, standard error of the mean (SEM)[40p.119-120], or Bland and Altman's limits of agreements[41], (3) G-theory is based on analysis of variance and further allows to partition the total score variance into all its variance components (main effects of each error sources and their interactions) which can be analyzed to make recommendations on optimal data-collection strategies to achieve desired levels of reliability.[42]

Gcoefs[43, p. 104] with Feldt 95% confidence intervals (95% CI)[43, p. 198-200] were computed by using the GENOVA[44] software program to estimate the reliability of each quantitative and qualitative measure at all disc levels. Significant differences in reliability between measures were identified by non-overlapping 95% CIs. The value of the Gcoef can vary between 0 and 1, with 1 indicating perfect reliability. Researchers have suggested that Gcoefs above 0.70 should be required for inferences about groups[45] and between 0.85 and 0.94 as a minimum to make inferences about an individual.[40, p.121]

The following repeated measures designs were used to estimate the reliability of quantitative measures: intra-evaluator for high-resolution images = 31 subjects X 3 segmentations; inter-evaluator for high-resolution images = 31 subjects X 2 evaluators segmenting images once; and inter-resolution = 31 subjects X 2 image resolutions X 1 segmentation per image by each of 2 evaluators. The design used to calculate the reliability for qualitative measures at high resolution was 31 subjects X 2 raters X 2 ratings per image.

The measurement error for relative decisions was also computed for each measure.[43, p. 102] This error is defined as the error of measurement when a person's score is interpreted *not* in terms of a standard of performance but, rather, in terms of group performance.[42, p. 84] The proportions of error variance associated with intra-, inter-evaluator, and inter-resolutions differences were calculated for quantitative measures and the proportions of error variance associated with intra- and inter-rater differences were calculated for qualitative ratings.

RESULTS

Means, standard deviations, and measurement errors for all quantitative distance and area measures are reported in Table 2.1 for each lumbar disc. The intra-evaluator Gcoefs for the discs between L1 and S1 ranged between 0.75 and 0.94 (median 0.91) for the posterior disc contour and between 0.66 and 0.95 (median 0.84) for the anterior disc contour (Table 2.2). The inter-evaluator Gcoefs varied between 0.70 and 0.95 (median 0.87) posteriorly and between 0.57 and 0.94 (median 0.82) anteriorly (Table 2.2).

Differences in Reliability between Quantitative Measures

No statistically significant differences in intra- or inter-evaluator reliability were found within posterior or anterior measures at each vertebral level (Table 2.2). For each of the measures of interest, no significant differences were found between the reliability at each of the five lumbar discs for both intra- and inter-evaluator reliability. No significant differences were found between intra- and inter-evaluator reliability. Within the posterior or anterior measures, no significant differences were found between distance and area measures. Most comparisons of the reliability of anterior and posterior measures were not significant. In nine of the ten significant differences, the reliability of an anterior measure of disc contour was lower than that of a posterior measure (Table 2.2). In only one instance the posterior measure was less reliable than the anterior measure. None of the measures using theoretical corners showed reliability significantly lower than a measure using real corners. Within anterior or posterior measures, no significant differences were found between measures using theoretical and real corners.

Sources of Errors

The percentage of the total error variance due to differences in repeated measures by the same evaluator varied between 0.0 and 12.4% (median = 3.0%). The error variance due to inter-evaluator differences ranged from 0.0 to 24.3% (median = 5.0%). The error variance due to differences between low- and high-resolution images varied between 0.0 and 9.7% (median= 2.0%). (See Appendix D, Table i-iv for variance components and percentages of total error variance at each disc level)

Subjective Ratings' Reliability

Mean ratings, standard deviation, and delta error of measurement for the qualitative ratings of bulging and herniation are presented in Table 2.3. The median Gcoefs for qualitative ratings by any rater on any occasion for the whole lumbar spine were as follows: 0.34 for anterior bulging, 0.48 for posterior bulging, 0.40 for anterior herniation, and 0.53 for posterior herniation (Table 2.4). Inter-rater differences accounted for most of the error median 17.5% and intra-rater differences for the least (median 0.0%). (See Appendix D, Table v-vi for variance components and percentages of total error variance at each disc level)

DISCUSSION

As hypothesized, the reliability of the quantitative distance and area measures of disc contour abnormality was superior to the reliability of the qualitative ratings of disc bulging on images typical of current MR practice. For secondary objectives, the reliabilities of the distance and area measures were not statistically different. Contrary to this study's hypothesis, no measures using theoretical corners were statistically significantly more reliable than measures using real corners, although a tendency was observed in this direction. Also, image resolution, for the two resolutions compared, was not a substantial a source of error as hypothesized for quantitative measures of disc contour. Overall reliability was more affected by intra-, inter-evaluator, and unidentified sources of error.

Difference in Reliability

To my knowledge, the reliability of different quantitative distances and areas measures of disc contour on sagittal images has never been compared within a single study. Reliability in this study was higher or comparable to that in previous studies. However, the external validity of previous studies is questionable due to methodological problems (sampling,[27,28] statistics used,[27] and dichotomizing of the measure[33]), which limit comparisons with this study. Furthermore, in this study, like Cooley et al.[28], we did not find statistically significant differences in the reliability of the quantitative measures between disc levels. Using the 95%CI may have been too stringent to identify differences between the reliability of different measures of different lumbar levels measures. However, until validity is examined, a measure should be excluded only if it shows a relatively important reliability problem.

The tendency for anterior measures to have lower reliability than posterior measures could be due to the following factors complicating segmentation: (1) the contrast between anterior blood vessels and the annulus does not always allow differentiation;[46] (2) the signal intensity of the anterior spine is lower;[47] (3) anterior osteophytes are more frequent;[48] and (4) the anterior longitudinal ligament is thicker, and longitudinal ligaments are difficult to differentiate from the annulus on MR images.[49,50]

Measures of disc contour using theoretical corners as defined in this study allow one to avoid uncertainty when segmenting the vertebral corners, which can be difficult to outline precisely.[25] We observed a tendency (borderline significance) for measures using theoretical corners to be more reliable than measures using real corners. The decision to select measures using real or theoretical corners cannot be made solely on the basis of reliability results. For instance, measures using real corners are easier to obtain.

Sources of Errors

Wood et al. found, in the thoracic spine, that the error in reading an image and retaking the image, with three repeat measures from two different evaluators, was 2.8%.[20] Although these results are similar to the median inter-evaluator error for a single segmentation (5%) found in this study, they are not directly comparable because they are from different anatomical levels (unspecified thoracic levels) in subjects with contour abnormalities diagnosed qualitatively.

Image resolution was not as important a source of error as expected. However, both evaluators reported working faster, feeling less fatigue when segmenting at true doubled resolution, and also agreed that artificially increasing resolution with a smoothing algorithm did not lead to those benefits. Higher image resolutions may help reduce measurement error. Alternatively, higher field strength could possibly allow an increase in resolution and reduction in slice thickness without compromising the contrast-to-noise ratio while reducing the potential for volume averaging artifacts.[51]

Other Sources of Error

Brightness and contrast levels adjustment when segmenting could also affect reliability. However, because signal intensity varies both vertically and horizontally within a sagittal image, standardizing those parameters when segmenting as Tonami et al.[22] have done on axial images where signal intensity varies less is difficult. Automatic segmentation algorithms could also be useful in reducing the intra- and inter-evaluator error.[52-54] However, accurate segmentation techniques for the disc are not yet available.

The outlines segmented on the proton density images were transposed onto other MR sequences and modified in order to increase the anatomical validity of the outlines. However, because of chemical shift artifact[50] and potential movement between imaging sequences, this procedure may have introduced some variability into the segmented outlines. Because of the vertical orientation of the frequency-encoding gradient in our system, the chemical shift artifact produces a thickened peripheral line of higher signal intensity along the interface between the water content of the disc and the fat content of the vertebral body below the disc and of higher signal intensity above the disc. This artifact affects the MR sequences with different TE to different degrees and may affect reliability. Using only one sequence may increase reliability at the expense of validity.

Reliability of Average Measures

The reliability that could be expected by averaging measures from a number of evaluators can be estimated by using generalizability theory and the data from this study.[55] Averaging a single segmentation by three different evaluators would allow quantitative measurement of disc contour with excellent reliability adequate for individual inferences (median expected Gcoefs 0.95 depending on the disc level). Such an approach, while time-consuming, could be useful in research. In contrast, averaging qualitative scores from three different raters would allow qualitative bulging ratings with reliability expected to be, at best, adequate for group inferences (median estimates based on data from the current study as between 0.61 and 0.73 depending on the disc level). Averages from multiple quantitative measures of disc contour have previously been used as the unit for analysis.[20,28,56]

Comparing Qualitative and Quantitative Measures

When using the same high-quality MR images, the reliability of quantitative measures was superior to that of experienced clinicians' qualitative ratings of disc bulging. The reliability of a single quantitative measure, but not that of qualitative measures, was adequate to make inferences at the group level. As in other studies of qualitative ratings,[8,57-59] the reliability of both the qualitative and quantitative ratings used in this study improved for subjects with more severe contour abnormalities (results not shown). The difference in reliability between quantitative and qualitative measures may have been smaller if the study sample had come from a patient population, including subjects with larger contour abnormalities. Nevertheless, this study's results suggest that quantitative measures are more appropriate for studies of population-based samples.

This study demonstrates that quantitative measures of disc contour abnormality show substantially higher reliability than qualitative ratings. The reliability of a single quantitative measure reached levels adequate for group comparisons for research purposes.[45] Quantitative measures are recommended for research into the determinants, pathogenesis, and significance of disc contour abnormalities. At this stage, quantitative measures are not recommended for routine clinical use because the clinical relevance of the various sizes of measured contour abnormalities is unknown for diagnostic or therapeutic decisions.

Table 2.1. Mean values, standard deviation (SD), and error of measurements for quantitative measures of disc contour abnormality for each level of the lumbar spine (n=31)

	Area measures[1]	**L1L2** **Mean** **(SD;delta)**	**L2L3** **Mean** **(SD;delta)**	**L3L4** **Mean** **(SD;delta)**	**L4L5** **Mean** **(SD;delta)**	**L5S1** **Mean** **(SD;delta)**
Posterior	Real corners	17.59[2] (6.56;2.21)[3]	23.23 (7.01;3.50)	30.53 (11.98;3.33)	35.84 (13.75;5.54)	24.05 (11.61;4.11)
	Theoretical corners	20.81 (6.77;2.65)	27.66 (6.95;3.32)	36.33 (11.68;2.77)	40.72 (13.05;4.95)	26.48 (12.29;4.00)
Anterior	Real corners	19.60 (11.02;4.58)	28.31 (12.91;5.40)	33.33 (12.89;7.89)	32.86 (14.91;8.54)	33.10 (15.05;5.16)
	Theoretical corners	34.64 (15.53;5.36)	45.74 (19.91;6.85)	50.67 (16.97;7.25)	56.12 (19.41;9.76)	74.33 (21.71;5.32)
	Distance measures[4]					
Posterior	Real corners	2.37 (0.74;0.34)	2.75 (0.81;0.39)	3.24 (1.20;0.36)	4.07 (1.27;0.48)	3.41 (1.11;0.66)
	Theoretical corners	2.89 (0.85;0.38)	3.43 (0.85;0.38)	4.01 (1.24;0.35)	4.56 (1.15;0.42)	3.65 (1.11;0.59)
Anterior	Real corners	2.19 (0.88;0.62)	2.59 (0.80;0.45)	2.79 (0.82;0.60)	2.46 (1.00;0.63)	2.50 (1.02;0.38)
	Theoretical corners	3.50 (1.39;0.59	3.89 (1.20;0.75)	3.92 (1.06;0.50)	3.96 (1.24;0.66)	5.50 (1.15;0.40)

[1] Area measures in millimeter-square, [2] Group mean of the average of the measures by the two evaluators at high resolution.
[3] Standard deviation, delta error of measurement for relative decisions[44], [4] Distance measures in millimeters

Table 2.2. Intra-, inter-evaluator generalizability coefficients, and 95% confidence intervals for the reliability for the quantitative measures of disc contour abnormality for each level of the lumbar spine (n=31)

		L1L2		L2L3		L3L4		L4L5		L5S1		L1S1	
	Areas In pixel^2	**Gcoeff Intra**^{1}	**Gcoef Inter**^{2}	**Gcoef Intra**	**Gcoef Inter**	**Gcoef Intra**	**Gcoef Inter**	**Gcoef Intra**	**Gcoef Inter**	**Gcoef Intra**	**Gcoef Inter**	**Gcoef Intra**	**Gcoef Inter**
Posterior	Real corners	.88 (.75;.94)	**.89**^{4} **(.79;.95)**	.83 (.68;.92)	**.78** **(.60;.89)**	.93 (.84;.97)	**.93**^{4} **(.84;.97)**	.93 (.84;.97)	**.85** **(.70;.93)**	.93 (.84;.97)	**.88** **(.76;.95)**	0.90	0.87
	Theoretical corners	.89 (.78;.95)	**.86** **(.72;.94)**	.88 (.75;.94)	**.80** **(.62;.90)**	.92 (.84;.97)	**.95**^{4} **(.88;.98)**	.91 (.80;.96)	**.87** **(.73;.94)**	.94 (.86;.97)	**.90** **(.79;.96)**	0.91	0.87
Anterior	Real corners	.68 (.47;.83)	**.84** **(.69;.93)**	.91 (.81;.96)	**.84** **(.69;.93)**	.79 (.61;.90)	**.68**^{3} **(.48;.83)**	.76 (.57;.88)	**.72** **(.52;.86)**	.89 (.77;.95)	**.89** **(.77;.95)**	0.81	0.79
	Theoretical corners	.88 (.75;.94)	**.89** **(.77;.95)**	.91 (.81;.96)	**.89** **(.77;.95)**	.85 (.71;.93)	**.83** **(.68;.92)**	.82 (.66;.92)	**.78** **(.59;.89)**	.95 (.89;.98)	**.94**^{4} **(.87;.97)**	0.88	0.87
	Distances In pixel												
Posterior	Real corners	.81 (.64;.91)	**.81** **(.64;.91)**	.75 (.56;.88)	**.79** **(.62;.90)**	.92 (.82;.96)	**.92**^{4} **(.82;.96)**	.92 (.84;.97)	**.87** **(.73;.94)**	.88 (.76;.95)	**.70**^{3} **(.50;.85)**	0.86	0.82
	Theoretical corners	.86 (.73;.94)	**.82** **(.66;.92)**	.90 (.79;.96)	**.82** **(.66;.92)**	.94^4 (.87;.97)	**.92**^{4} **(.84;.97)**	.92 (.82;.96)	**.88** **(.75;.94)**	.88 (.76;.95)	**.75** **(.56;.88)**	0.90	0.84
Anterior	Real corners	.66 (.45;.82)	**.60**^{3} **(.39;.78)**	.78 (.60;.90)	**.73** **(.53;.87)**	.74^3 (.55;.86)	**.57**^{3} **(.36;.76)**	.78 (.59;.89)	**.67** **(.46;.83)**	.84 (.69;.93)	**.87** **(.74;.94)**	0.76	0.69
	Theoretical corners	.84 (.69;.93)	**.83** **(.67;.92)**	.83 (.68;.92)	**.67** **(.46;.83)**	.77 (.58;.89)	**.80** **(.62;.90)**	.85 (.71;.93)	**.75** **(.56;.88)**	.92 (.82;.96)	**.89** **(.76;.95)**	0.84	0.79

[1] Intra = Evaluator (EP) manually segmented images three times at two resolutions,

[2] Inter= first manual segmentation by both evaluators at high resolution

[3] Significantly smaller than another anterior or posterior measure at the same level.

[4] Significantly larger than another anterior or posterior measure at the same level

Table 2.3. Mean values, standard deviation (SD), and delta error of measurements for qualitative measures of disc bulging/herniation for each level of the lumbar spine

Variables	L1L2 Mean (SD; delta)	L2L3 Mean (SD; delta)	L3L4 Mean (SD; delta)	L4L5 Mean (SD; delta)	L5S1 Mean (SD; delta)
Anterior bulging[1]	.50[2] (1.12;.83)[3]	.80 (1.33;1.01)	.69 (1.23;.65)	.79 (1.30;1.04)	.23 (.63;.34)
Posterior bulging[1]	.45 (.93;.51)	.85 (1.21;0.78)	1.65 (1.52;1.18)	2.12 (1.54;0.95)	1.60 (1.57;1.23)
Anterior herniation[1]	.18 (.37;.08)	.32 (.52;.15)	.27 (.44;.10)	.27 (.41;.11)	.10 (.26;.06)
Posterior herniation[1]	.17 (.28;.05)	.30 (.34;.06)	.57 (.42;.09)	.73 (.47;.10)	.55 (.42;.07)

n=31, from resolution study.

[1] Bulging and herniation were assessed qualitatively by using a 4-point rating scale.

[2] Group mean of the average of both measures by raters 1 and 2 on high-resolution images.

[3] Standard deviation, delta error of measurement for relative decisions[44]

Table 2.4. Generalizability coefficients estimates and 95% confidence intervals of the reliability for the qualitative measures of disc bulging/herniation on high-resolution images by any rater on any occasion for each level of the lumbar spine

	L1L2 Gcoef	L2L3 Gcoef	L3L4 Gcoef	L4L5 Gcoef	L5S1 Gcoef	L1S1 Gcoef
	.29	.38	.52	.34	.00[3]	
Anterior bulging[1]	(.18;.47)[2]	(.24;.56)	(.36;.70)	(.21;.52)	(.00;.00)	.31
	.41	.47	.48	.63[4]	.51[4]	
Posterior bulging[1]	(.27;.60)	(.32;.66)	(.33;.67)	(.47;.78)	(.35;.69)	.50
	.41	.40	.47	.27[3]	.00[3]	
Anterior herniation[1]	(.26;.59)	(.26;.59)	(.32;.66)	(.16;.45)	(.00;.00)	.31
	.43	.52	.53	.57	.61[4]	
Posterior herniation[1]	(.28;.62)	(.36;.70)	(.37;.70)	(.41;.74)	(.45;.77)	.53

n=31 from resolution study,

[1] Bulging and herniation were assessed qualitatively by using a 4-points rating scale.

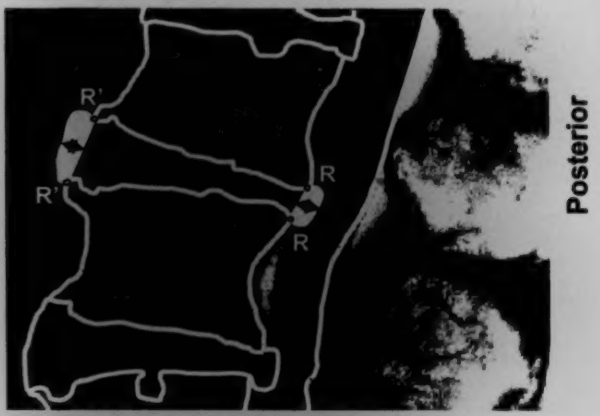
[2] Generalizability coefficient (95% confidence interval) from two repeated ratings by two raters [44]

[3] Significantly smaller than another measure at the same level.

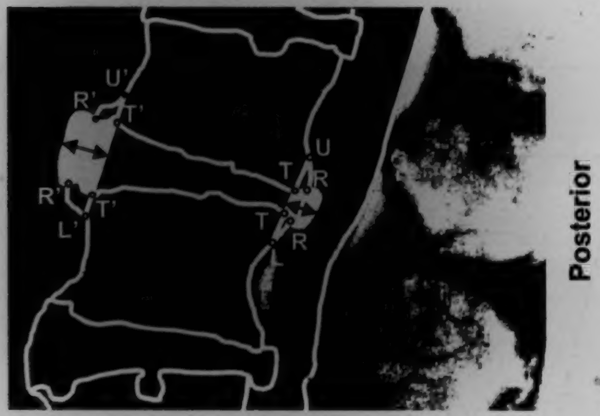
[4] Significantly larger than another anterior or posterior measure at the same level

Figure 2.1. Measures of disc contour abnormality, using real and theoretical corners

Using Real Corners



Using Theoretical Corners



Top: Measures of anterior and posterior area of disc contour abnormality, using real corners (R) (White surface) and measures of anterior and posterior distance of disc contour abnormality, using real corners as references (R) (double headed arrows).

Bottom: Measures of anterior and posterior area of disc contour abnormality, using theoretical corners (T) (White surface) and measures of anterior and posterior distance of disc contour abnormality, using theoretical corners as references (T) (double headed arrows). Theoretical corners are defined by the intersections (T) of a line connecting the diameter of the vertebra above the disc at 20% of vertebra height (U) to the corresponding diameter below the disc (L) with a line representing the disc-vertebra boundary.

REFERENCES – CHAPTER 2

1. Mixter WJ, Barr JS. Rupture of the intervertebral disc with involvement of the spinal canal. *N Engl J Med.* 1934;1934(211):210-215.
2. Raininko R, Manninen H, Battie MC et al. Observer variability in the assessment of disc degeneration on magnetic resonance images of the lumbar and thoracic spine. *Spine.* 1995;20(9):1029-1035.
3. Parkkola R, Rytokoski U, Kormano M. Magnetic resonance imaging of the discs and trunk muscles in patients with chronic low back pain and healthy control subjects. *Spine.* 1993;18(7):830-836.
4. Jarvik JG, Haynor DR, Koepsell TD et al. Interreader reliability for a new classification of lumbar disk disease. *Acad Radiol.* 1996;3(7):537-544.
5. Brant-Zawadzki MN, Jensen MC, Obuchowski N et al. Interobserver and intraobserver variability in interpretation of lumbar disc abnormalities. A comparison of two nomenclatures. *Spine.* 1995;20(11):1257-1263.
6. Boos N, Semmer N, Elfering A et al. Natural history of individuals with asymptomatic disc abnormalities in magnetic resonance imaging: predictors of low back pain-related medical consultation and work incapacity. *Spine.* 2000;25(12):1484-1492.
7. Karppinen J, Malmivaara A, Tervonen O et al. Severity of symptoms and signs in relation to magnetic resonance imaging findings among sciatic patients. *Spine.* 2001;26(7):E149-E154.
8. Milette PC, Fontaine S, Lepanto L et al. Differentiating lumbar disc protrusions, disc bulges, and discs with normal contour but abnormal signal intensity. Magnetic resonance imaging with discographic correlations. *Spine.* 1999;24(1):44-53.
9. Jarvik JJ, Hollingworth W, Heagerty P et al. The longitudinal assessment of imaging and disability of the back (laidback) study: baseline data. *Spine.* 2001;26(10):1158-1166.
10. Battie MC, Videman T, Gibbons LE et al. 1995 Volvo Award in clinical sciences. Determinants of lumbar disc degeneration. A study relating lifetime exposures and magnetic resonance imaging findings in identical twins. *Spine.* 1995;20(24):2601-2612.
11. Boos N, Rieder R, Schade V et al. 1995 Volvo Award in clinical sciences. The diagnostic accuracy of magnetic resonance imaging, work perception, and psychosocial factors in identifying symptomatic disc herniations. *Spine.* 1995;20(24):2613-2625.
12. Glass GV, Hopkins KD. *Statistical methods in education and psychology.* 3 ed. Needham Heights, MA, USA: Allyn and Bacon, 1995,1-674.
13. Videman T, Leppavuori J, Kaprio J et al. Intragenic polymorphisms of the vitamin D receptor gene associated with intervertebral disc degeneration. *Spine.* 1998;23(23):2477-2485.
14. Videman T, Nummi P, Battie MC et al. Digital assessment of MRI for lumbar disc desiccation. A comparison of digital versus subjective assessments and digital intensity profiles versus discogram and macroanatomic findings. *Spine.* 1994;19(2):192-198.

15. Bird P, Ejbjerg B, McQueen F et al. OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Studies. Exercise 5: an international multicenter reliability study using computerized MRI erosion volume measurements. *J Rheumatol.* 2003;30(6):1380-1384.
16. Botsford DJ, Esses SI, Ogilvie-Harris DJ. In vivo diurnal variation in intervertebral disc volume and morphology. *Spine.* 1994;19(8):935-940.
17. Bozzao A, Gallucci M, Masciocchi C et al. Lumbar disk herniation: MR imaging assessment of natural history in patients treated without surgery. *Radiology.* 1992;185(1):135-141.
18. Holodny AI, Kisza PS, Contractor S et al. Does a herniated nucleus pulposus contribute significantly to a decrease in height of the intervertebral disc? Quantitative volumetric MRI. *Neuroradiology.* 2000;42(6):451-454.
19. Kato F, Mimatsu K, Kawakami N et al. Serial changes observed by magnetic resonance imaging in the intervertebral disc after chemonucleolysis. A consideration of the mechanism of chemonucleolysis. *Spine.* 1992;17(8):934-939.
20. Wood KB, Blair JM, Aepple DM et al. The natural history of asymptomatic thoracic disc herniations. *Spine.* 1997;22(5):525-529.
21. Buttermann GR. Lumbar disc herniation regression after successful epidural steroid injection. *J Spinal Disord Tech.* 2002;15(6):469-476.
22. Tonami H, Yokota H, Nakagawa T et al. Percutaneous laser discectomy: MR findings within the first 24 hours after treatment and their relationship to clinical outcome. *Clin Radiol.* 1997;52(12):938-944.
23. Ebeling U, Reulen HJ. Are there typical localisations of lumbar disc herniations? A prospective study. *Acta Neurochir (Wien).* 1992;117(3-4):143-148.
24. Fries JW, Abodeely DA, Vijungco JG et al. Computed tomography of herniated and extruded nucleus pulposus. *J Comput Assist Tomogr.* 1982;6(5):874-887.
25. Vincent JM, Baldwin JE, Sims C et al. Vertebral 'corner' defect associated with lumbar disk herniation shown by magnetic resonance imaging. *Spine.* 1993;18(1):109-113.
26. Luoma K, Riihimaki H, Raininko R et al. Lumbar disc degeneration in relation to occupation. *Scand J Work Environ Health.* 1998;24(5):358-366.
27. Fredericson M, Lee SU, Welsh J et al. Changes in posterior disc bulging and intervertebral foraminal size associated with flexion-extension movement: a comparison between L4-5 and L5-S1 levels in normal subjects. *Spine J.* 2001;1(1):10-17.
28. Cooley JR, Danielson CD, Schultz GD et al. Posterior disk displacement: morphologic assessment and measurement reliability-lumbar spine. *J Manipulative Physiol Ther.* 2001;24(5):317-326.
29. Sahlstrand T, Lonntoft M. A prospective study of preoperative and postoperative sequential magnetic resonance imaging and early clinical outcome in automated percutaneous lumbar discectomy. *J Spinal Disord.* 1999;12(5):368-374.
30. Matsubara Y, Kato F, Mimatsu K et al. Serial changes on MRI in lumbar disc herniations treated conservatively. *Neuroradiology.* 1995;37(5):378-383.
31. Luoma K, Riihimaki H, Luukkonen R et al. Low back pain in relation to lumbar disc degeneration. *Spine.* 2000;25(4):487-492.

32. Silcox DH, III, Horton WC, Silverstein AM. MRI of lumbar intervertebral discs. Diurnal variations in signal intensities. *Spine.* 1995;20(7):807-811.
33. Luoma K, Vehmas T, Raininko R et al. Lumbosacral transitional vertebra: relation to disc degeneration and low back pain. *Spine.* 2004;29(2):200-205.
34. Harada A, Okuizumi H, Miyagi N et al. Correlation between bone mineral density and intervertebral disc degeneration. *Spine.* 1998;23(8):857-861.
35. Videman T, Leskinen T, Stalhammar H et al. Effect of glycoaminoglycan polysulphate on the human intervertebral discs based on stature measurements. *Clin Biomech.* 1994;9(3):206-207.
36. Fardon DF, Milette PC. Nomenclature and classification of lumbar disc pathology. Recommendations of the Combined task Forces of the North American Spine Society, American Society of Spine Radiology, and American Society of Neuroradiology. *Spine.* 2001;26(5):E93-E113.
37. Stratford PW, Norman GR, McIntosh JM. Generalizability of grip strength measurements in patients with tennis elbow. *Phys Ther.* 1989;69(4):276-281.
38. Crocker L, Algina J. *Introduction to classical & modern test theory.* First ed. Orlando, Florida: Harcourt Brace Jovanovich College Publishers, 2002,1-482.
39. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin.* 1979;86(2):420-428.
40. Streiner DL, Norman GR. *Health measurement scales: A practical guide to their development and use.* Second ed. New York, New York: Oxford Medical Publications, 1995,1-231.
41. Bland JM, Altman GG. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research.* 1999;8(2):135-160.
42. Shavelson RJ, Webb NM. *Generalizability Theory -- A Primer.* First ed. Newbury Park, California: Sage Publications, 1991,1-136.
43. Brennan RL. *Generalizability theory.* New York: Springer, 2001,1-538.
44. Crick JE, Brennan RL. *GENOVA: A GENeralized Analysis Of VAriance System* [computer program]. Version 2.1, Iowa City, Iowa: The American College Testing Program, 1983.
45. Scientific Advisory Committee of the Medical Outcome Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res.* 2002;11(3):193-205.
46. Wiltse LL, Fonseca AS, Amster J et al. Relationship of the dura, Hofmann's ligaments, Batson's plexus, and a fibrovascular membrane lying on the posterior surface of the vertebral bodies and attaching to the deep layer of the posterior longitudinal ligament. An anatomical, radiologic, and clinical study. *Spine.* 1993;18(8):1030-1043.
47. Brault JS, Driscoll DM, Laakso LL et al. Quantification of lumbar intradiscal deformation during flexion and extension, by mathematical analysis of magnetic resonance imaging pixel intensity profiles. *Spine.* 1997;22(18):2066-2072.
48. Hilton RC, Ball J. Vertebral rim lesions in the dorsolumbar spine. *Ann Rheum Dis.* 1984;43(2):302-307.
49. Silverman CS, Lenchik L, Shimkin PM et al. The value of MR in differentiating subligamentous from supraligamentous lumbar disk herniations. *AJNR Am J Neuroradiol.* 1995;16(3):571-579.

50. Grenier N, Greselle JF, Vital JM et al. Normal and disrupted lumbar longitudinal ligaments: correlative MR and anatomic study. *Radiology.* 1989;171(1):197-205.
51. Maubon AJ, Ferru JM, Berger V et al. Effect of field strength on MR images: comparison of the same subject at 0.5, 1.0, and 1.5 T. *Radiographics.* 1999;19(4):1057-1067.
52. Ramirez L, Parent E, Durdle NG et al. Interactive Segmentation for Quantitative Analysis of Disc and Vertebral Morphometric Characteristics". *International Research Society for Spinal Deformities.* Paper 45 . 2004.
53. Hoad CL, Martel AL. Segmentation of MR images for computer-assisted surgery of the lumbar spine. *Phys Med Biol.* 2002;47(19):3503-3517.
54. Roberts N, Gratin C, Whitehouse GH. MRI analysis of lumbar intervertebral disc height in young and older populations. *J Magn Reson Imaging.* 1997;7(5):880-886.
55. Shavelson RJ, Webb NM. Generalizability and Decision Studies with Different Designs. In: Foster D, editor. *Generalizability Theory A Primer.* Newbury Park, California: SAGE Publications, Inc., 1991: 115-126.
56. Nakano M, Matsui H, Ishihara H et al. Serial changes of herniated intervertebral discs after posterior lumbar discectomy: the relation between magnetic resonance imaging of the postoperative intervertebral discs and clinical outcome. *J Spinal Disord.* 2001;14(4):293-300.
57. Beattie PF, Meyers SP, Stratford P et al. Associations between patient report of symptoms and anatomic impairment visible on lumbar magnetic resonance imaging. *Spine.* 2000;25(7):819-828.
58. Boden SD, Davis DO, Dina TS et al. Abnormal magnetic-resonance scans of the lumbar spine in asymptomatic subjects. A prospective investigation. *J Bone Joint Surg Am.* 1990;72(3):403-408.
59. Pfirrmann CW, Dora C, Schmid MR et al. MR image-based grading of lumbar nerve root compromise due to disk herniation: reliability study with surgical correlation. *Radiology.* 2004;230(2):583-588.

CHAPTER 3: MEASUREMENT ERROR DUE TO INTRA-EVALUATOR AND INTER-SCANNER DIFFERENCES IN QUANTITATIVE MEASURES OF LUMBAR DISC CONTOUR ABNORMALITY ON MR IMAGES

INTRODUCTION

Longitudinal studies are increasingly needed to improve our knowledge of the etiology and pathogenesis of disc contour abnormalities. As well, mounting evidence suggests that reductions in the dimensions of disc contour abnormalities occur as a result of natural history and interventions.[1-8] Reliable measures will help reduce the sample size needed to detect the effects of natural history or interventions on disc bulging or herniation,[6,8] which are referred to as "disc contour abnormalities" in this paper.

To address the need for precise measurements, quantitative measures of disc contour abnormality on magnetic resonance (MR) images are being developed and have been shown to be more reliable than medical specialists' traditional qualitative ratings.[9,10] Quantitative measures are ideal for follow-up and correlation etiologic studies because they have more measurement levels than qualitative rating scales, thereby increasing the likelihood of detecting change.[11]

Factors influencing the reliability of quantitative measures of disc contour abnormality from MR images studied to date include intra-,[9,10,12-15] inter-evaluator,[9,10,14,15] and image-resolution variations.[10] However, to my knowledge, the majority of reliability studies of quantitative disc contour abnormality measurements have been conducted using the same images,[9,10,12,13] inadequate sampling,[9] or inadequate statistics.[15] Only two studies, with inadequate sample size (n=1), have estimated the scan-rescan reliability of disc measures and found it to be lower than intra- or inter-evaluator reliability.[14,15]

Differences due to re-imaging subjects in different scanners are also likely to influence the reliability of quantitative measures of disc contour abnormality. Studies of the influence of inter-scanner differences on measurement properties for the lumbar spine are rare.[16-18] The scanners tested in previous studies had different field strengths, and the measures tested were either qualitative[17,18] or were not direct measures of disc morphology.[16] Furthermore, in a study estimating the reliability of measures from a 1.5 Tesla (T) MR scanner and a 0.3T scanner, different subjects were imaged on both scanners.[18] Thus, the important question of whether differences between scanners with the same field strength influence the reliability of measures obtained from the same subjects has not been answered yet. The comparability of repeated images is a particularly important issue for follow-up studies, as the lifetime of a MR scanner rarely exceeds 10 years and software upgrades, influencing image acquisition, are likely to happen within three to five years, a typical duration for a longitudinal study.[19] Inter-scanner differences are also important when interpreting the results of multi-center studies where different MR scanners are used.[19]

The impact of inter-scanner differences on quantitative measures of segmented areas on MR images has been estimated in brain and cardio-vascular imaging studies.[20-22] Inter-scanner error (7.5 -18.5%) was generally larger than scan-rescan error (1.9 - 4.8%) or intra-evaluator error (1.8 - 4.2%).[20-22] Thus, inter-scanner differences are expected to influence the reliability of quantitative measures of contour abnormality, possibly more so than intra-evaluator differences.

The objectives of the present study were (1) to determine the reliability of quantitative measures of lumbar disc contour abnormality on MR images obtained on different MR scanners, (2) to determine the proportion of measurement error due to differences between MR scanners (including subject repositioning that occurs in test-retest situations) and between repeated measures from a single evaluator, and (3) to compare the reliability of distance and area measures of disc contour abnormality.

MATERIALS AND METHODS

Subjects

Thirty-four male subjects from a population-based sample participating in a larger epidemiological study[23] were enrolled between April 1997 and February 1998. These subjects were not selected on the basis of LBP history or the presence or absence of clinical findings. The mean (± standard deviation) subject age was 48.5 (± 5.9) years. During the 12 months prior to the evaluation, the subjects reported a mean of 5.5 (± 11.0) days of difficulty doing daily work due to low back problems. The majority of subjects had disc contour abnormalities extending less than six millimeters posteriorly, a magnitude corresponding to a moderate posterior disc bulging assessed qualitatively (L1L2 = 91%, L2L3 = 73.5%, L3L4 = 61.8%, L4L5 = 50%, L5S1 =41.2%). All subjects received written information about the study and provided written consent before participation.

Imaging

Lumbar spine images were obtained for all subjects using two MR scanners on normal maintenance routines: a 1.5T Siemens Magnetom SP4000 scanner (Numaris software A2.7) and a Magnetom Vision scanner (Numaris software B31b) (Siemens AG, Germany). The subjects were imaged lying supine with a pillow supporting their knees. Subjects were imaged in one or the other scanner first, depending on availability, then walked to the other scanner where a different technician conducted the imaging protocol. Before obtaining the images of interest, subjects spent approximately 30 minutes in the supine position to allow for most of the effect of compression due to weight-bearing on the disc dimensions to normalize.[24] Proton density (PD), T1-, and T2-weighted midsagittal lumbar images (T12-S1) were obtained by using a surface coil and a matrix size of 256 X 256 (pixel size 1.02 mm). Two excitations were averaged, and the phase encoding direction was anterior to posterior. The repetition time (TR) and echo time (TE) for each sequence were selected by a radiologist to maximize the similarity between the Vision and the SP4000 images (SP4000: PD (TR/TE = 2600/22), T2-(2600/90), and T1-(650/22) weighted; Vision: PD (TR/TE = 2600/22), T2-(2600/90), and T1-(650/12)). Other differences between the MR scanners included the gradient strength (SP 4000 was 10 mT/m, and Vision was 25 mT/m) and the coils (SP 4000 had a single loop coil, and Vision used a 4-element phased array coil).

Quantitative Assessment

Custom-designed spine image analysis software (SpEx© v2.42) was used to obtain quantitative measures of disc contour abnormality. The contours of the lumbar discs, vertebrae, and spinal canal were segmented manually on the midsagittal PD image by a single evaluator taking 15-20 minutes per subject (Figure 3.1). The evaluator selected the image where the spinous processes of L4 and L5 vertebra were the most clearly revealed. The evaluator segmented vertically the vertebral wall's contour from the longitudinal ligament. To segment the contour of the disc annulus, at the endplate level, the evaluator moved through the longitudinal ligament to follow the outer contour of the ligament/intervertebral disc. The contrast did not allow the annulus to be distinguished from the ligament, so the ligament was included in the disc area. To segment the disc from the vertebra, the evaluator followed horizontally the boundary between the vertebral bone and the endplates. Cartilaginous endplates were included in the disc since the contrast did not allow both tissues to be distinguished. The areas created by the intersection of those segmentation lines formed the regions of interests corresponding to the discs and vertebrae from which the measures were derived (Figure 3.1).[10,25-27] The segmented areas were then adjusted by using T2- (disc) and T1-weighted (vertebra) images, if necessary, taking advantage of different contrasts.

The distances and areas of disc contour abnormality were calculated by using the segmented outlines from two types of reference points (Figure 3.1). Disc contour abnormality was measured by using the "real" vertebral corners (Figure 3.1: R and R') as seen on the MR image at the intersection of the vertebral wall and the endplate contours. The second type of disc contour abnormality measures used the "theoretical" corners (Figure 3.1: T and T') of the vertebra as a reference. "Theoretical corners" were defined by the intersections of a line connecting the diameter of the vertebra above the disc at 20% of the vertebra height to the corresponding diameter below the disc, with the lines representing the disc-vertebra boundaries (Figure 3.1). Measures using theoretical corners were developed to address the difficulty of precisely identifying the position of the real corners on MR images with low contrast and to measure disc contour abnormality when combined with the presence of osteophytes. In a previous study, the median intrarater reliability generalizability coefficients calculated for repeated quantitative measures of disc contour abnormality, using this procedure on the same 512X512 image, were 0.84 (range from 0.66 to 0.95) anteriorly and 0.91 (range from 0.75 to 0.94) posteriorly.[10,26] Such levels are adequate for group-level inferences.[28]

To estimate the intra-rater and inter-scanner reliability of quantitative disc contour abnormality measures, the evaluator, with prior experience manually segmenting over 300 spine MR images, segmented all the images. Segmentation was completed twice for each image from both scanners (SP4000 first, Vision second) at least a month apart to estimate both intra-evaluator and inter-scanner errors. The evaluator was blind to his other segmentations and to the participants' medical histories and identities.

Analyses

Generalizability coefficients (Gcoefs)[29, p. 104] with Feldt 95% confidence intervals (95% CI)[29, p. 198-200] were computed using the GENOVA[30] software program to estimate the reliability of each measure at all disc levels. Significant differences in reliability between measures were identified by non-overlapping 95%CI. The following repeated measures design was used to estimate the Gcoefs: 34 subjects X 2 scanners X 2 segmentations by the same evaluator. The value of the Gcoef can vary between 0 and 1, with 1 indicating perfect reliability. Researchers have suggested that Gcoefs above 0.70 be required to make inferences about groups[28] and that reliability coefficients between 0.85 and 0.94 as a minimum be required to make inferences for an individual.[31, p.121] The measurement error for relative decisions was also computed for each measure.[29, p. 102] This error is defined as the error of measurement when a person's score is interpreted *not* in terms of a standard of performance but, rather, in terms of group performance.[32, p. 84] Generalizability theory allows the estimation of the proportion of the total error variance that can be attributed to each of the sources of error studied. Thus, the error variance components associated with inter-scanner and intra-evaluator differences were also calculated. The average Gcoefs were obtained by using Fisher Z transformation to avoid underestimation of the mean due to the skewness of the distribution.[33, p. 362] The averages of the measurement errors between the disc levels were calculated as the square root of the average of the error variances.

RESULTS

The mean dimensions of disc contour abnormality and the standard deviation from the four measurements (two segmentations on images from two scanners) obtained from all subjects in this study are presented in Table 3.1.

Reliability Coefficients

The generalizability coefficients for the reliability of disc contour abnormality measures by a single evaluator on any occasion, using images from any MR scanner, varied between 0.68 and 0.89 (mean 0.82) for the posterior measures, and between 0.53 and 0.89 (mean 0.77) for the anterior measures, depending on the disc level (Table 3.2). The generalizability coefficients of 24/25 posterior measures and 20/25 anterior measures were above 0.70.

Difference in Reliability between Disc Levels

For all posterior measures and those anterior measures using theoretical corners, no statistically significant differences were found among disc levels. Measuring anterior contour abnormality using real corners was less reliable at L3L4 (distance and area) and at L1L2 (area) than at L5S1.

Difference in Reliability between Measures at a Given Disc Level

Anterior vs posterior: For most comparisons, the reliability of measures of anterior disc contour abnormality did not differ significantly from that of corresponding posterior measures. However, four statistically significant differences in reliability were found. In three of these cases, the anterior measures were less reliable than the posterior measures. At level L1L2, the distance of disc contour abnormality measured by using theoretical corners was more reliable anteriorly than posteriorly. At L2L3 (area, using theoretical corners) and L3L4 (distance and area, using real corners), disc contour abnormality was more reliably measured posteriorly.

Within posterior or anterior measures: None of the posterior disc contour abnormality measures had an estimated reliability significantly larger than any other at a given lumbar disc level. Statistically significant differences in reliability amongst the anterior disc contour abnormality measures were found at only two levels. At L1L2, distance and area measures using theoretical corners were more reliable than measures using real corners ($p < 0.05$). At L3L4, the area measure of contour abnormality, using theoretical corners, was more reliable than the distance measure of anterior contour abnormality, using real corners. No clear pattern emerged to suggest that the reliability of either area or distance was better.

Measurement Errors

The mean measurement error across the disc levels for quantitative measurement of disc contour abnormality, using real corners on any given MR scanner on any occasion by a single evaluator, was 18 mm^2 (error/mean measure = 22%) for posterior areas, 40 mm^2 (29%) for anterior areas, 1 mm (20%) for posterior distance measures, and 1.4 mm (25%) for anterior distance measures (Table 3.1). For quantitative measures using theoretical corners, the mean measurement error was 19 mm^2 (19%) for posterior areas, 46 mm^2 (19%) for anterior areas, 1 mm (16%) for posterior distance measures, and 1.4 mm (15%) for anterior distance measures (Table 3.1).

Partitioning Error Variance

The percent error variance (error variance/total variance) due to inter-scanner differences varied between 2.1% and 21.4% (mean 8.3%) and that due to intra-evaluator differences varied between 0.0% and 5.8% (mean 1.2%) (Table 3.3). Similar portions of total error variance explained were observed for each lumbar disc level. Also, similar partitioning was observed for anterior and posterior measures, and for measures using real or theoretical corners. The majority of the error variance for both the area and distance measurements remained unexplained. (See Appendix E, Table i to ii for variance components and percentage of error error variance at each disc level)

DISCUSSION

We found that quantitative measures of disc contour abnormality obtained from images where the subjects are repositioned in different MR scanners have adequate reliability for group level inference (> 0.70).[28] The measurement error observed was more influenced by the different MR scanners than intra-evaluator variations. The median error of measurement across levels for one measure on an image from any MR scanner corresponded to approximately 20% of the mean measures of disc contour abnormality obtained in this study's sample.

The inter-scanner error estimate includes error due to variations in subject positioning and scanner properties, and slight differences in sequence parameters inherent in comparisons of images obtained from different scanners. The two scanners used in this study were manufactured by the same company. Scanners from different companies could lead to larger inter-scanner differences. Alternatively, improvements in technology since the data in this study were collected may reduce the differences between scanners. However, an attempt had been made to reduce the inter-scanner measurement error by choosing imaging parameters maximizing image similarity between scanners, yet some differences were still noted. Currently, improvements in imaging sequences, coils, and shielding could still lead to differences between scanners large enough to warrant consideration in longitudinal or multicenter studies.

Error estimates were obtained only for the midsagittal slice in the current study. The same digitizing method could be applied on lateral slices and may lead to similar reliability estimates. However, because the Gcoefs tended to be larger for measures or levels with larger mean contour abnormality (graph not shown), the reliability could increase if lateral slices, where larger contour abnormalities are more prevalent, were analyzed.[34] Alternatively, reliability may decrease if the presence of pedicles or the contrast between structures of interest is not as good as on the midsagittal images, which may create additional challenges for the manual segmentation procedure. Thus, reliability should be assessed in other planes to verify these hypotheses..

This study included an adequate number of subjects representative of a range of disc degeneration. Although the sample included many participants who would not seek care from medical specialists, approximately half the participants had contour abnormalities larger than six millimeters at levels L4L5 and L5S1. This finding is consistent those of with previous studies demonstrating a high prevalence of contour abnormalities in subjects with and without symptoms.[35-39] The results are likely generalizable to subjects with more severe findings as well because the plots of reliability estimates against the mean contour abnormality measures (not shown) demonstrated that reliability increased with measurements of more severe contour abnormalities. Previous studies of inter-scanner reliability in other areas of research have often included a smaller number of subjects and/or restricted their samples to healthy subjects, limiting the generalizability of the results.

Comparisons with the Existing Literature

To my knowledge, the inter-scanner reliability of the quantitative measures of disc contour abnormality has not been studied. The reliability observed in this study was comparable or slightly lower (mean difference = approximately 0.07) than in my previous study of the reliability of the quantitative measures obtained repeatedly from the same images. This finding is not surprising because the images used in this study had lower resolution and were obtained on older scanners producing noisier images. Indeed, the previous study showed that using images with higher resolution (pixel size = 0.5 mm rather than 1 mm) had a small positive effect (2%) on the reliability of quantitative measures of disc contour abnormality.[10] In that study, the percent error due to intra-evaluator differences was 3%, and that due to inter-evaluator differences was 5%.[10] The findings from the present study suggest that inter-scanner differences are of a slightly greater magnitude (8.3%) than those found earlier for inter-evaluator variations.

Few significant differences in reliability between the quantitative measures of disc contour abnormality were found in this study, but these differences suggested that the reliability may be better for measures using theoretical corners and measures of posterior disc contour. This observation is consistent with that of my previous reliability study.[10] No clear pattern of differences was found between lumbar levels or when comparing areas to distance measures.

The external validity of most other studies of quantitative measures' reliability using repeated measures of the same images is questionable due to methodological problems that limit comparisons with this study. These include sampling,[9,15] statistics used,[15] and dichotomizing of the measure.[12] The ratio of the dimension of contour abnormality to the spinal canal area on axial images has been assessed,[13,14,40-42] and the inter-evaluator error was estimated to be as low as 2.8% for an average of three segmentations,[13,14] which is comparable to my earlier findings on sagittal images.[10] Other sagittal distance measures of disc contour abnormality[9,15,43] and ratios of contour abnormalities to spinal canal[8,40,41,44] have been used. The reliability estimates reported for such measures vary widely (ICC: 0.40 to 0.92).[9,12] The reliability of the measures described in this study is at the higher end of the range of others' reliability estimates when obtained from the same images and slightly below the upper end when obtained from different scanners.[10]

Implications and Recommendations

Differences in disc contour abnormality larger than the measurement error (approximately 20% in this study) should be observed before being confident that the differences measured are not due to chance alone and are potentially clinically meaningful.[19] This recommendation affects sample size calculations in planning cross-sectional or longitudinal studies. For example, using data for posterior measures of disc contour abnormality at L4L5 (Table 3.1), we can illustrate the impact on sample size estimation. In cross-sectional studies, 71 to 99 subjects would be needed in each group in order to detect a difference between groups imaged on different MR scanners, in L4L5 posterior disc contour abnormality corresponding to the magnitude of the measurement error with a power of 0.80 at an alpha level of 0.05.[45] In a longitudinal study, if the same group of subjects were imaged on different MR scanners at baseline and follow-up, 38 to 51 subjects would be needed to detect a difference the size of the "measurement error" in quantitative disc contour abnormality (with power = 0.80 and $\alpha = 0.05$).[45] Such samples are realistic when one considers the prevalence of contour abnormalities in the population[46] or in persons with back pain.[36]

Measurement error corresponded to approximately 20% of the mean disc contour abnormalities observed in this study. Changes in disc contour abnormality larger than 20% have been reported over periods varying between 10 weeks to three years of conservative treatment by using qualitative measures or quantitative measures with unknown reliability.[3,5-7,47-49] These studies support the adequacy of the measures described in this paper for comparisons of disc contour abnormalities in cross-sectional and longitudinal studies, even when different MR scanners set to produce comparable images are used.

In conclusion, the differences between measurements obtained from images acquired on different MR scanners have more influence on the reliability of the measures of disc contour abnormality than intra-evaluator differences. The intra-evaluator/inter-scanner reliability of the quantitative measures of disc contour abnormality described in this study is adequate for group-level inferences.[28] However, a clinically important difference in disc contour abnormality should be at least as large as the measurement error, which corresponds to approximately 20% of the mean disc contour abnormalities measured in this study. The subjects' position, MRI parameters, MR scanner variations, and image evaluation procedures should be carefully controlled in follow-up and multi-center studies of disc contour abnormalities.

Table 3.1. Mean values, standard deviation (SD), and error of measurement for quantitative measures of disc contour abnormality by any evaluator on any scanner for each level of the lumbar spine

	Area measures (mm^2)	L1L2 Mean (SD; error)	L2L3 Mean (SD; error)	L3L4 Mean (SD; error)	L4L5 Mean (SD; error)	L5S1 Mean (SD; error)
Posterior	Real	55.0[1] (23.9; 14.3)[2]	70.9 (35.6; 14.4)	85.1 (38.3; 16.0)	102.6 (44.0; 20.7)	95.8 (57.0; 24.4)
	Theoretical	69.3 (28.5;14.9)	90.1 (40.7; 13.8)	104.8 (44.4; 17.2)	118.7 (46.3; 21.7)	109.1 (67.1; 23.7)
Anterior	Real	103.7 (63.3; 42.4)	134.4 (72.8; 46.2)	132.2 (51.1; 35.8)	159.9 (69.5; 37.6)	159.3 (95.8; 37.1)
	Theoretical	171.4 (95.9; 37.0)	212.6 (89.9; 52.9)	211.7 (72.0; 36.8)	275.8 (96.5; 47.2)	325.1 (130.4; 51.1)
	Distance measures (mm)					
Posterior	Real	3.7 (1.5; 1.0)	4.5 (2.0; 1.0)	5.0 (1.9; 0.9)	6.4 (2.4; 1.1)	6.4 (2.7; 1.2)
	Theoretical	4.9 (1.9; 1.1)	5.9 (2.4; 0.9)	6.4 (2.2; 1.0)	7.4 (2.5; 1.0)	6.9 (2.9; 1.1)
Anterior	Real	5.3 (2.8; 1.8)	5.9 (2.3; 1.4)	5.5 (1.6; 1.3)	5.9 (2.3; 1.2)	5.8 (2.5; 1.4)
	Theoretical	8.1 (4.3; 1.5)	8.7 (2.8; 1.5)	8.1 (2.1; 1.2)	9.6 (2.7; 1.4)	11.8 (3.2; 1.3)

[1] Mean of reading 1 and 2 on Magnetom SP4000 and Vision imagers, [2] Standard deviation; error of measurement for relative decision for measurement by any evaluators on any imager[30]

Table 3.2. Generalizability coefficients (Gcoef) estimates and 95% confidence intervals of the reliability for the quantitative measures of disc contour abnormality by any evaluator on any MR scanner for each level of the lumbar spine

	Area measures	L1L2 Gcoef (upper; lower limit)	L2L3 Gcoef (upper; lower limit)	L3L4 Gcoef (upper; lower limit)	L4L5 Gcoef (upper; lower limit)	L5S1 Gcoef (upper; lower limit)	L1S1 Mean Gcoef
Posterior	Real	.71 (.59; .81)	.85 (.77; .91)	.84[2] (.76; .91)	.81 (.71; .88)	.84 (.75; .90)	0.81
	Theoretical	.77 (.66; .86)	.89[2] (.83; .94)	.86 (.79; .92)	.81 (.71; .88)	.89 (.82; .93)	0.84
Anterior	Real	.65[1] (.53; .78)	.68 (.56; .80)	.63[1] (.50; .75)	.75 (.64; .85)	.86 (.79; .92)	0.72
	Theoretical	.86[2] (.79; .92)	.72[1] (.60; .82)	.78[2] (.68; .87)	.79 (.69; .87)	.86 (.78; .92)	0.80
	Distance measures						
Posterior	Real	.68 (.56; .80)	.80 (.71; .88)	.80[2] (.71; .88)	.82 (.73; .89)	.83 (.74; .90)	0.79
	Theoretical	.71[1] (.60; .81)	.86 (.79; .92)	.82 (.73; .89)	.86 (.79; .92)	.88 (.82; .93)	0.83
Anterior	Real	.66[1] (.53; .78)	.71 (.59; .82)	.53[1] (.40; .67)	.79 (.69; .87)	.75 (.64; .84)	0.69
	Theoretical	.89[2] (.82; .93)	.75 (.64; .85)	.74 (.63; .84)	.77 (.67; .86)	.85 (.77; .91)	0.80

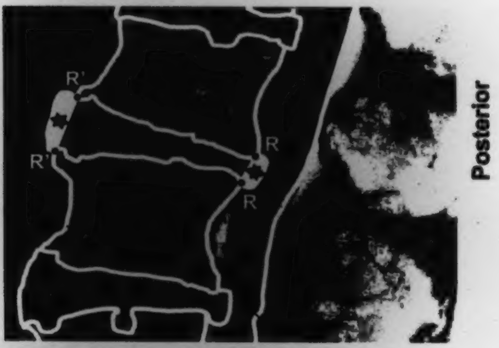
[1] Significantly smaller than another anterior or posterior measure at the same level.

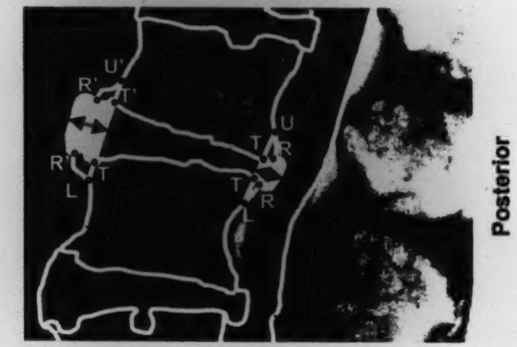
[2] Significantly larger than another anterior or posterior measure at the same level.

Table 3.3. Percent error variance due to inter-scanner and intra-evaluator differences at each lumbar disc level

		L1L2		L2L3		L3L4		L4L5		L5S1		L1S1	
Areas		% Inter-scanner	% Intra-evaluator	% Inter-scanner	% Intra-evaluator	% Inter-scanner	% Intra-evaluator	% Inter-scanner	% Intra-evaluator	% Inter-scanner	% Intra-evaluator	% Inter-scanner	% Intra-evaluator
Posterior	Real	15.0	0.0	5.5	2.3	8.7	0.6	7.8	2.6	3.9	2.1	8.2	1.5
	Theoretical	14.2	0.0	3.7	1.2	6.6	1.0	10.4	1.5	2.7	0.7	7.5	0.9
Anterior	Real	12.6	0.0	11.5	0.0	12.8	0.0	10.5	0.0	6.4	0.0	10.8	0.0
	Theoretical	4.2	0.0	8.5	0.0	4.0	0.0	9.8	0.0	7.6	0.0	6.8	0.0
Distances													
Posterior	Real	12.1	0.0	7.0	5.8	8.3	1.4	8.0	3.4	6.1	1.3	8.3	2.4
	Theoretical	16.1	5.6	4.8	3.3	8.1	2.9	6.2	0.9	2.1	0.0	7.5	2.5
Anterior	Real	13.9	2.9	5.9	0.0	21.4	4.4	4.9	0.0	9.1	0.0	11.0	1.5
	Theoretical	4.6	0.9	5.1	0.0	8.3	0.0	7.5	1.7	5.0	0.0	6.1	0.5

Figure 3.1. Measures of disc contour abnormality, using real and theoretical corners


Using Real Corners


Using Theoretical Corners

Top: Measures of anterior and posterior area of disc contour abnormality, using real corners (R) (White surface), and measures of anterior and posterior distance of disc contour abnormality, using real corners as references (R) (double headed arrows). Real corners (R) are defined by the intersection of two lines: the vertical outline of the vertebrae following the bone contour continuing with the outside contour of the discs with the horizontal outlines of the contour of the endplates.

Bottom: Measures of anterior and posterior area of disc contour abnormality, using theoretical corners (T) (White surface), and measures of anterior and posterior distance of disc contour abnormality, using theoretical corners as references (T) (double headed arrows). Theoretical corners are defined by the intersections (T) of a line connecting the diameter of the vertebra above the disc at 20% of vertebra height (U) to the corresponding diameter below the disc (L) with a line representing the disc-vertebra boundary.

REFERENCES – CHAPTER 3

1. Delauche-Cavallier MC, Budet C, Laredo JD et al. Lumbar disc herniation. Computed tomography scan changes after conservative treatment of nerve root compression. *Spine.* 1992;17(8):927-933.
2. Teplick JG, Haskin ME. Spontaneous regression of herniated nucleus pulposus. *AJR Am J Roentgenol.* 1985;145(2):371-375.
3. Komori H, Shinomiya K, Nakai O et al. The natural history of herniated nucleus pulposus with radiculopathy. *Spine.* 1996;21(2):225-229.
4. Saal JA, Saal JS, Herzog RJ. The natural history of lumbar intervertebral disc extrusions treated nonoperatively. *Spine.* 1990;15(7):683-686.
5. Ellenberg MR, Ross ML, Honet JC et al. Prospective evaluation of the course of disc herniations in patients with proven radiculopathy. *Arch Phys Med Rehabil.* 1993;74(1):3-8.
6. Gallucci M, Bozzao A, Orlandi B et al. Does postcontrast MR enhancement in lumbar disk herniation have prognostic value? *J Comput Assist Tomogr.* 1995;19(1):34-38.
7. Kawaji Y, Uchiyama S, Yagi E. Three-dimensional evaluation of lumbar disc hernia and prediction of absorption by enhanced MRI. *J Orthop Sci.* 2001;6(6):498-502.
8. Matsubara Y, Kato F, Mimatsu K et al. Serial changes on MRI in lumbar disc herniations treated conservatively. *Neuroradiology.* 1995;37(5):378-383.
9. Cooley JR, Danielson CD, Schultz GD et al. Posterior disk displacement: morphologic assessment and measurement reliability-lumbar spine. *J Manipulative Physiol Ther.* 2001;24(5):317-326.
10. Parent E, Videman T, Rogers T et al. The influence of MRI image quality on the reliability of quantitative measures of disc bulging/herniation. *5th Combined Meeting of the Orthopaedic Research Societies of Canada, U.S.A., Japan and Europe.* 310. 10-10-2004.
11. Bird P, Ejbjerg B, McQueen F et al. OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Studies. Exercise 5: an international multicenter reliability study using computerized MRI erosion volume measurements. *J Rheumatol.* 2003;30(6):1380-1384.
12. Luoma K, Vehmas T, Raininko R et al. Lumbosacral transitional vertebra: relation to disc degeneration and low back pain. *Spine.* 2004;29(2):200-205.
13. Buttermann GR. Lumbar disc herniation regression after successful epidural steroid injection. *J Spinal Disord Tech.* 2002;15(6):469-476.
14. Wood KB, Blair JM, Aepple DM et al. The natural history of asymptomatic thoracic disc herniations. *Spine.* 1997;22(5):525-529.
15. Fredericson M, Lee SU, Welsh J et al. Changes in posterior disc bulging and intervertebral foraminal size associated with flexion-extension movement: a comparison between L4-5 and L5-S1 levels in normal subjects. *Spine J.* 2001;1(1):10-17.
16. Greenstein J, Cramer GD, Howe J et al. Comparison of 1.5 Tesla and 0.35 Tesla field strength magnetic resonance imaging scans in the morphometric evaluation of

the lumbar intervertebral foramina. *J Manipulative Physiol Ther.* 1995;18(4):195-202.

17. Merl T, Scholz M, Gerhardt P et al. Results of a prospective multicenter study for evaluation of the diagnostic quality of an open whole-body low-field MRI unit. A comparison with high-field MRI measured by the applicable gold standard. *Eur J Radiol.* 1999;30(1):43-53.
18. Cihangiroglu M, Yildirim H, Bozgeyik Z et al. Observer variability based on the strength of MR scanners in the assessment of lumbar degenerative disc disease. *Eur J Radiol.* 2004;51(3):202-208.
19. Tofts PS. Standardisation and optimisation of magnetic resonance techniques for multicentre studies. *J Neurol Neurosurg Psychiatry.* 1998;64 Suppl 1:S37-S43.
20. Westwood MA, Anderson LJ, Firmin DN et al. Interscanner reproducibility of cardiovascular magnetic resonance T2* measurements of tissue iron in thalassemia. *J Magn Reson Imaging.* 2003;18(5):616-620.
21. Filippi M, van Waesberghe JH, Horsfield MA et al. Interscanner variation in brain MRI lesion load measurements in MS: implications for clinical trials. *Neurology.* 1997;49(2):371-377.
22. Filippi M, Rocca MA, Gasperini C et al. Interscanner variation in brain MR lesion load measurements in multiple sclerosis using conventional spin-echo, rapid relaxation-enhanced, and fast-FLAIR sequences. *AJNR Am J Neuroradiol.* 1999;20(1):133-137.
23. Simonen R, Levalahti E, Kaprio J et al. Multivariate Genetic Analysis of Lifetime Exercise and Environmental Factors. *Med Sci Sports Exerc.* 2004;36(9):1559-1566.
24. Videman T, Leskinen T, Stalhammar H et al. Effect of glycoaminoglycan polysulphate on the human intervertebral discs based on stature measurements. *Clin Biomech.* 1994;9(3):206-207.
25. Videman T, Nummi P, Battie MC et al. Digital assessment of MRI for lumbar disc desiccation. A comparison of digital versus subjective assessments and digital intensity profiles versus discogram and macroanatomic findings. *Spine.* 1994;19(2):192-198.
26. Parent E, Battie MC, Videman T. Quantitative MRI Measures of Disc Bulging/Herniation: Development and Reliability. *First Annual Bone and Joint Annual Meeting.* 1[1]. 10-24-2003.
27. Battie MC, Videman T, Gibbons LE et al. 1995 Volvo Award in clinical sciences. Determinants of lumbar disc degeneration. A study relating lifetime exposures and magnetic resonance imaging findings in identical twins. *Spine.* 1995;20(24):2601-2612.
28. Scientific Advisory Committee of the Medical Outcome Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res.* 2002;11(3):193-205.
29. Brennan RL. *Generalizability theory.* New York: Springer, 2001,1-538.
30. Crick JE, Brennan RL. *GENOVA: A GENeralized Analysis Of VAriance System* [computer program]. Version 2.1, Iowa City, Iowa: The American College Testing Program, 1983.

31. Streiner DL, Norman GR. *Health measurement scales: A practical guide to their development and use.* Second ed. New York, New York: Oxford Medical Publications, 1995,1-231.
32. Shavelson RJ, Webb NM. *Generalizability Theory -- A Primer.* First ed. Newbury Park, California: Sage Publications, 1991,1-136.
33. Glass GV, Hopkins KD. *Statistical methods in education and psychology.* 3 ed. Needham Heights, MA, USA: Allyn and Bacon, 1995,1-674.
34. Ebeling U, Reulen HJ. Are there typical localisations of lumbar disc herniations? A prospective study. *Acta Neurochir (Wien).* 1992;117(3-4):143-148.
35. Jensen MC, Brant-Zawadzki MN, Obuchowski N et al. Magnetic resonance imaging of the lumbar spine in people without back pain. *N Engl J Med.* 1994;331(2):69-73.
36. Boos N, Rieder R, Schade V et al. 1995 Volvo Award in clinical sciences. The diagnostic accuracy of magnetic resonance imaging, work perception, and psychosocial factors in identifying symptomatic disc herniations. *Spine.* 1995;20(24):2613-2625.
37. Boden SD, Davis DO, Dina TS et al. Abnormal magnetic-resonance scans of the lumbar spine in asymptomatic subjects. A prospective investigation. *J Bone Joint Surg Am.* 1990;72(3):403-408.
38. Burns JW, Loecker TH, Fischer JR, Jr. et al. Prevalence and significance of spinal disc abnormalities in an asymptomatic acceleration subject panel. *Aviat Space Environ Med.* 1996;67(9):849-853.
39. Weinreb JC, Wolbarsht LB, Cohen JM et al. Prevalence of lumbosacral intervertebral disk abnormalities on MR images in pregnant and asymptomatic nonpregnant women. *Radiology.* 1989;170(1 Pt 1):125-128.
40. Bozzao A, Gallucci M, Masciocchi C et al. Lumbar disk herniation: MR imaging assessment of natural history in patients treated without surgery. *Radiology.* 1992;185(1):135-141.
41. Kato F, Mimatsu K, Kawakami N et al. Serial changes observed by magnetic resonance imaging in the intervertebral disc after chemonucleolysis. A consideration of the mechanism of chemonucleolysis. *Spine.* 1992;17(8):934-939.
42. Tonami H, Yokota H, Nakagawa T et al. Percutaneous laser discectomy: MR findings within the first 24 hours after treatment and their relationship to clinical outcome. *Clin Radiol.* 1997;52(12):938-944.
43. Luoma K, Riihimaki H, Raininko R et al. Lumbar disc degeneration in relation to occupation. *Scand J Work Environ Health.* 1998;24(5):358-366.
44. Sahlstrand T, Lonntoft M. A prospective study of preoperative and postoperative sequential magnetic resonance imaging and early clinical outcome in automated percutaneous lumbar discectomy. *J Spinal Disord.* 1999;12(5):368-374.
45. Dupont WD, Plummer WD, Jr. Power and sample size calculations. A review and computer program. *Control Clin Trials.* 1990;11(2):116-128.
46. Battie MC, Videman T, Parent E. Lumbar Disc Degeneration: Epidemiology and Genetic Influences. *Spine.* 2004;29(23):2679-2690.
47. Ahn SH, Park HW, Byun WM et al. Comparison of clinical outcomes and natural morphologic changes between sequestered and large central extruded disc herniations. *Yonsei Med J.* 2002;43(3):283-290.

48. Ahn SH, Ahn MW, Byun WM. Effect of the transligamentous extension of lumbar disc herniations on their regression and the clinical outcome of sciatica. *Spine.* 2000;25(4):475-480.

49. Nakano M, Matsui H, Ishihara H et al. Serial changes of herniated intervertebral discs after posterior lumbar discectomy: the relation between magnetic resonance imaging of the postoperative intervertebral discs and clinical outcome. *J Spinal Disord.* 2001;14(4):293-300.

CHAPTER 4: THE EFFECT OF LUMBAR FLEXION AND EXTENSION ON DISC CONTOUR ABNORMALITY MEASURED QUANTITATIVELY ON MRI

INTRODUCTION

Mixter and Barr postulated that herniated discs were the origin of lumbar spine symptoms in 1934, and today, the pathology of the lumbar intervertebral disc continues to be a suspected cause of low back pain (LBP).[1-7] Some researchers have suggested that disc contour abnormalities (bulging and herniation) can cause LBP by mechanically, and/or chemically irritating the outer layers of the annulus,[2,8] or other innervated structures within the spinal canal.[2,9-11]

Magnetic resonance (MR) imaging of the lumbar spine is considered the examination of choice for diagnosing intervertebral disc pathology.[12,13] The assessment of disc contour abnormalities has traditionally been done on MR images by using qualitative ratings.[14,15] However, the reliability of qualitative ratings of contour abnormalities has been shown to be suboptimal.[16-21] The intra- and inter-rater reliability of quantitative measures of disc contour abnormality on MR images is superior to that of qualitative ratings.[22,23]

Clinically, MR images to assess the lumbar spine are obtained in the neutral spine position. Radiographic or MR images in flexion, extension, or rotation have primarily been used to investigate instability.[24-27] However, the assessment of degenerative findings in the lumbar spine has shown positional dependence. Cadaver studies,[28-32] *in vivo* myelography,[33,34] discography,[35] conventional MR,[36-40] and, more recently, open-field MR[24,34,38,41-43] have been used to demonstrate the positional dependence of measurements of spinal canal stenosis,[24,32,34,36,38,39,43] intervertebral foramen stenosis,[24,28,29,32,36,38,39,41,43] nerve root compromise,[30,32,41] nucleus position,[35,37,40] and disc height.[36,40] The positional dependence of disc contour abnormality has been studied *in vitro*,[29,31,32] indirectly by using myelography,[33] qualitatively by using open MR scanners,[24,41,43] and in a quantitative open MR study with inadequate sample size (n = 3).[42] Although authors have suggested that differences in the dimensions of disc contour abnormalities between positions could help explain differences in nerve root compromise, spinal canal, or foramen stenosis,[29,32,36,38,39,41,43] very little attention has been devoted to quantitatively measuring *in vivo* differences in contour abnormality between positions. The extent to which variations in positioning affect the apparent size of disc contour abnormality within the bore of the most widely used conventional MR scanners is unclear. Knowledge of the effect of positioning on lumbar disc contour abnormality would be useful in research and clinical situations where disc contour abnormality is being evaluated.

The primary objective of this study was to determine the effects of flexing and extending the lumbar spine, on the size of disc contour abnormality assessed quantitatively. More specifically, we (1) examined the size of lumbar disc contour abnormality measured quantitatively in subjects with the spine in flexed, neutral, and extended positions, and (2) determined whether large disc contour abnormalities were associated with larger changes in size between positions.

MATERIAL AND METHODS

Subjects

Male volunteers over 21 years old were recruited by using bulletin board notices. Potential subjects were not selected on the basis of LBP history, and subjects of different ages were included to ensure the representation of a range of disc degeneration. The modified Oswestry Low Back Disability Questionnaire[44] and a pain history questionnaire were used to describe the subjects' clinical status. The study protocol was approved by the Ethics Review Committee of the investigators' institution. All subjects received written information about the study and provided written consent before participation.

Twenty-six males between 24 and 74 years old (median 40 years) volunteered. Oswestry disability levels were between 0 and 36 (median 4) out of a possible score of 100 indicating maximum disability. Pain on the day of the evaluation varied between 0 and 4 (median 0) on an 11-point numeric scale. Fifty-four percent of the subjects had had an episode of back problems that had limited their ability to work or participate in leisure activities.

Imaging

Lumbar MR images (L1-S1) were obtained with the participant supine with the spine in the neutral position (Figure 4.1a), in maximal flexion possible within the bore of the 1.5T Siemens Sonata imager (Figure 4.1b), and in extension (Figure 4.1c). Images in all positions were obtained in a single imaging session in this order: neutral, flexion, and extension. Subjects were comfortably supported in each of the positions in order to prevent movement artifacts. Before obtaining the images of interest, subjects spent approximately 30 minutes in the supine position to allow for most of the effect of compression due to weight-bearing on the disc dimensions to normalize.[45] For the neutral and flexion images, lumbar spine array coils were used. For the extension images, a body coil was used to ensure a sufficient signal was recorded. T1-weighted turbo spin echo mid-sagittal images in each position were obtained with repetition time (TR) of 650 ms and echo time (TE) of 24 ms. Two excitations were used. Slice thickness was 4 mm. The full phase acquisition matrix was 256 x 256, the field of view was 260 mm, and the pixel size was 1.02 mm. The image acquisition time was 5 minutes, 55 seconds.

Quantitative Measures of Disc Contour Abnormality

Custom-designed spine image analysis software (SpEx© v2.42) was used to obtain quantitative measures of disc contour abnormality. The contours of the lumbar discs, vertebrae, and spinal canal were segmented manually on the midsagittal T1 image in each position by an experienced evaluator, who selected the midsagittal image where the spinous processes of L4 and L5 vertebrae were the most clearly revealed (Figure 4.2). The contour of the vertebral wall was vertically segmented from the longitudinal ligament. To segment the contour of the disc annulus, at the endplate level, the evaluator moved through the longitudinal ligament to follow the outer contour of the ligament/intervertebral disc, since the contrast did not allow the annulus to be distinguished from the ligament. To segment the disc from the vertebra, the boundary between the vertebral bone and the endplate was followed. The areas created by the intersections of the segmentation lines formed the regions of interest corresponding to the discs and vertebrae from which measures were derived (Figure 4.2).[17,23,46,47]

Distances and areas were then calculated by using the segmented outlines (Figure 4.2). Disc contour abnormality was measured using the vertebral corners (Figure 4.2: V and V') as seen on the MR image at the intersection of the vertebral wall and the endplate contours. The median intra-rater reliability coefficients for the quantitative measures of disc contour abnormality were 0.84 anteriorly and 0.91 posteriorly.[23,47] Such levels are adequate for group-level inferences.[48] The angle between the line connecting the corners of the endplates above and below a disc represented the segmental angle. All manual segmentations for each position were completed at least two weeks apart. The evaluators were blind to the participants' identities and medical histories.

Quantitative measures derived from the segmentation of the images obtained in extension were corrected by a factor calculated from the difference in the dimension of the vertebral body (used as internal control) between images obtained with and without the body coil and the lumbar roll. Such a correction is needed because of the difference in sensitivity of the spine array coil for the extension image where there is more distance between the spine and the coil array when placing a lumbar roll under L3. The correction factor was calculated as an average of the differences in the midsagittal area of the vertebrae above and below the disc between the extension image and the flexion as well as the difference with the neutral image. The analysis of the vertebral body dimension showed that there was no need for correction between the flexion and neutral images.

Analyses

Analysis of variance for repeated measures was used to test for differences in segmental angles and quantitative disc measures among flexed, neutral and, extended positions. Least significant difference (LSD) pairwise tests were used for post-hoc comparisons. The percent difference between positions was calculated by using the mean contour abnormality in the neutral position as the denominator (e.g., % difference in flexion vs. neutral position = [(contour abnormality in flexion – contour abnormality in neutral position) / contour abnormality in neutral position] * 100). Pearson's coefficient was used to study correlations.

RESULTS

Angles

The disc segmental angles at each motion segment were reduced in flexion (difference Δ 2.3-2.9 degrees, $p < 0.05$) and increased in extension compared to the neutral position (Δ 1.4-4.7 degrees, $p < 0.05$), except at L5S1 between the flexion and neutral positions (Table 4.1).

Dimensions of Posterior Disc Contour Abnormality

The posterior contour abnormality was smaller in flexion (Δ 15.2-27.4%, $p<0.05$) and in extension (Δ 9.5-30.1%, $p < 0.05$) than in the neutral position at all disc levels (Table 4.2). The posterior contour abnormality was most often not statistically significantly different between flexed and extended spines. (Table 4.2).

Dimensions of Anterior Disc Contour Abnormality

Anterior disc contour abnormality was larger in flexion than in extension (Δ 17.2-71.7%, $p<0.05$) for all disc levels (Table 4.3). Also, anterior contour abnormality measures in the neutral position were typically intermediary. Anterior contour abnormality was larger in flexion than in the neutral position at all disc levels when using distance measures (Δ 26.3-52.7%, $p < 0.05$) and only at L3L4 and L5S1 when measuring the area of contour abnormality (Δ 25.0% and 19.8%, $p < 0.05$). Anterior contour abnormality was larger in the neutral position than in extension for all levels except L3L4 (Δ 16.0-31.1%, $p < 0.05$) (Table 4.3).

Correlation between Difference in Quantitative Contour Abnormality between Positions and Contour Abnormality Severity in the Neutral Position

Posterior Disc Contour Abnormality

Greater posterior disc contour abnormality in neutral was associated with greater decreases in the dimensions of posterior contour abnormality when moving from the neutral position to flexion (9/10 correlations $p < 0.05$, except for the area at L5S1, Figure 4.3a and 4.4a) or extension (9/10 correlations $p < 0.05$, except L3L4 for the area, Figure 4.3b and 4.4b). The magnitude of contour abnormality in the neutral position explained a mean of 54% and 34% of the variance in the differences in contour abnormality dimensions between the flexed and neutral positions and between the extended and neutral positions, respectively.

Anterior Disc Contour Abnormality

Greater anterior disc contour abnormality in the neutral position was associated with larger decreases in the dimensions of anterior contour abnormality when moving from flexion to the neutral position (7/10 correlations, $p < 0.05$ at all levels except at L5S1 depending on the measure) and from the neutral position to extension (10/10 correlations, $p < 0.05$). For statistically significant correlations, the magnitude of contour abnormalities in the neutral position explained a mean of 55% and 54% of the variance in the differences in contour abnormalities between the flexed and the neutral position and between the neutral position and extension, respectively.

DISCUSSION

Positioning in the sagittal plane does influence quantitative MR measures of disc contour abnormality. In particular, MR imaging of the lumbar spine in a neutral supine position consistently yielded the largest posterior contour abnormality measurements, as compared to positions emphasizing spine flexion or extension in the bore of the MR scanner. Furthermore, to my knowledge, this study is the first to describe the positional dependence of the size of anterior disc contour abnormality *in vivo*. While anterior disc findings are less likely to influence neurological symptoms, anterior bulges could be a source of LBP by compression or stretching of pain-sensitive tissue.[2] The largest anterior contour abnormality measurements were observed on images in the flexed position. These findings suggest that, when evaluators are interested in documenting the degree of contour abnormality, MR images should be obtained with standardized positioning of the spine. This recommendation particularly applies for subjects with more severe contour abnormalities, who were found to have the largest differences in the dimensions of contour abnormalities between positions.

The Effect of Flexion and Extension on Disc Contour Abnormality

The results from this study are consistent with previous findings of smaller posterior contour abnormality in flexion.[29,32,41,43] However, in contrast to this study's finding that posterior contour abnormality decreased in extension compared to the neutral position, other studies found no change or showed increases in posterior contour abnormality.[29,32,41,43] Weishaupt et al. found that in most subjects posterior disc contour abnormality did not differ between positions when imaging subjects sitting in an open MR scanner and assessing posterior disc contour abnormality qualitatively.[41] Only, 5-8% of the discs had smaller contour abnormality in flexion than in the neutral position, in the neutral position than in extension, and in flexion than in extension.[41] Further, even less, 1.3% of the discs, had smaller contour abnormality in the neutral position than in extension.[41] Using a similar methodology, Zamani et al. found that posterior disc contour abnormality was smaller in flexion than in the neutral position in 30% of subjects, larger in extension than in the neutral position in 33.3%, and smaller in extension than in the neutral position in only 3.3%.[43] On CT images, the dimensions of posterior disc contour abnormality in the foramen decreased significantly in flexion, and increased, or did not differ significantly in extension.[29,32]

The finding that posterior contour abnormality decreased rather than increased when measured in extension when compared to the neutral position could have resulted because the range of movement available in this study was limited to that allowable within the bore of the MR scanner and because the spine was not loaded, as it was in previous cadaver studies and open MR examinations.[29,32,41,43] Furthermore, only the midsagittal image was quantitatively measured in the current study while other MR studies qualitatively assessed axial images, where para-sagittal lateral disc contour abnormalities could influence measurements.[41,43] These potential explanations are supported by *in vitro* findings where disc pressurization, spine position, and location on the disc (lateral vs. midline) have been shown to explain 56% of the variation in measured bulges[31] and by studies showing that axial compression increases disc contour abnormality.[31,42,49,50] However, loading the lumbar spine in flexion or extension within a conventional MR scanner is difficult. Nevertheless, this study's results are arguably the most relevant to the effects of position on the assessment of disc contour abnormality from MR as routinely obtained in most clinical settings.

The current study's results do not support the other authors' hypothesis that increases in posterior disc contour abnormality are responsible for the increase in the severity of stenosis for both the spinal canal and lateral foramen.[39,41-43] However, either contour abnormality was not specifically measured quantitatively in those studies[39,41,43] or the sample size was too small to draw a conclusion.[42] Other factors, such as thickening of the ligamentum flava, should be considered to explain the increase in stenosis observed in extension.[38]

Studies using quantitative measures of other vertebral findings [38,39,42] also found a greater number of significant differences than studies using qualitative measures, suggesting that quantitative measures are more sensitive to change.[34,41] This greater sensitivity could explain why the current study found significant changes in most comparisons while no change was apparent in a majority of cases in other studies using qualitative measures of posterior contour abnormality.[41,43]

Furthermore, reductions in the dimensions of posterior disc contour abnormality as large as 35% were observed in this study, which is more than the 20% reported over periods varying between 10 weeks to three years of conservative treatment using qualitative measures or quantitative measures with unknown reliability.[3,10,51-55] Thus, the variations in posterior disc contour abnormality quantified in this study are substantial.

Degree of Degenerative Findings and Differences between Positions

Our finding that differences in the magnitude of posterior disc contour abnormalities between positions are larger in subjects with more severe contour abnormalities agrees with observations in the literature.[41,43] Zamani et al. observed that 27% of all discs and 40% of degenerated discs studied had increased contour abnormality in extension.[43] Alternatively, the differences in the size of disc contour abnormality among positions could also correlate with other degenerative findings not examined in this study. Back pain patients with a higher Thomson grade of disc degeneration (reduced height and signal intensity) have more evidence of instability on flexion and extension images.[26,56] This finding suggests that, especially in subjects with degeneration, positioning for MR imaging should be standardized.

Study Strengths and Limitations

Among the strengths of this study is the adequate sample size used, which was larger than that in most previous studies.[27,38-40,42] A 1.5T MR scanner was used, which allows a stronger signal-to-noise ratio than most systems used in positional dependence studies reported to date.[24,34,38,40-43] Also, previous *in vivo* studies either did not measure anterior findings or did not use reliable quantitative measures of disc contour abnormality.[24,33,39,41-43] Unlike most other studies of disc contour abnormality, this study verified that positioning of the subjects to exaggerate flexion or extension actually led to differences in disc segmental angles to assist in interpreting differences in disc contour abnormality.[24,33,41-43]

This study does have limitations. None of the subjects had LBP for which they were currently seeking care from a medical specialist. This factor may reduce the strength of the correlations studied due to range restriction in the severity of disc contour abnormality and limit the generalizability of the results to patient populations. Measures of disc contour abnormality were obtained only from midsagittal images. However, if a posterior bulge increase were observed on midsagittal images obtained in the neutral position compared to extension, then an increase would also be expected on para-sagittal images. Nevertheless, such a hypothesis should be verified. The goal of the current study was to describe differences between positions and not to study the mechanisms leading to such differences. Furthermore, positions were held for only small amounts of time. Thus, this study did not address the effect of repeated or sustained end-range flexion or extension exercises.

In summary, both flexed and extended positioning of the spine, possible in a conventional MR scanner, have the potential to reduce the apparent size of posterior disc contour abnormality on MR images as compared to imaging in a neutral position. Exaggerated lumbar extension to the degree allowed in a standard MR scanner reduced the degree of disc contour abnormality both anteriorly and posteriorly. Thus, patient positioning should be standardized when an accurate estimate of the dimension of a disc contour abnormality is needed.

Table 4.1. Mean disc segmental angle in the flexed, neutral, and extended positions

Disc	Flexion*	Neutral	Extension	Overall p-value
L1L2	2.5 ± 2.7a	5.4 ± 2.8	10.1 ± 3.0 ac	< 0.001
L2L3	3.7 ± 2.4a	6.4 ± 3.8	10.5 ± 2.8 ac	< 0.001
L3L4	6.5 ± 4.6a	8.8 ± 4.4	11.6 ± 3.4 ac	< 0.001
L4L5	9.6 ± 4.3a	12.2 ± 3.7	13.6 ± 3.7 ac	< 0.001
L5S1	14.5 ± 4.7	15.9 ± 4.8	18.5 ± 5.5 ac	< 0.001

* Values are mean ± standard deviation in degrees.

Abbreviations: a = significantly different from the neutral position,

c = significant difference between the flexed and extended positions.($p < 0.05$)

Table 4.2. Mean quantitative posterior disc contour abnormality in the flexed, neutral, and extended positions

Disc	Flexion	Neutral	Extension	Overall p
	Posterior contour abnormality area (mm^2)			
L1L2	14.5 ± 5.6 a	19.3 ± 8.3	13.9 ± 5.9 a	< 0.001
L2L3	18.4 ± 6.4 a	25.4 ± 8.3	17.8 ± 5.6 a	< 0.001
L3L4	27.6 ± 11.9 a	35.4 ± 13.8	29.5 ± 13.9 a	< 0.001
L4L5	32.9 ± 12.1 a	41.5 ± 12.2	29.1 ± 11.2 ac	< 0.001
L5S1	20.9 ± 13.1 a	25.3 ± 13.1	20.1 ± 12.3 a	0.009
	Posterior contour abnormality distance (mm)			
L1L2	1.9 ± 0.6 a	2.5 ± 0.9	2.1 ± 0.8 a	< 0.001
L2L3	2.2 ± 0.7 a	3.1 ± 0.9	2.5 ± 0.7 ac	< 0.001
L3L4	3.0 ± 1.2 a	3.7 ± 1.4	3.3 ± 1.3 a	< 0.001
L4L5	3.6 ± 1.1 a	4.6 ± 1.2	3.5 ± 1.1 a	< 0.001
L5S1	3.1 ± 1.3 a	3.6 ± 1.5	3.0 ± 1.3 a	0.005

* Values are mean ± standard deviation in degrees.

Abbreviations: a = significantly different from the neutral position,

c = significant difference between the flexed and extended positions.($p< 0.05$)

Table 4.3. Mean quantitative anterior disc contour abnormality in the flexed, neutral, and extended positions

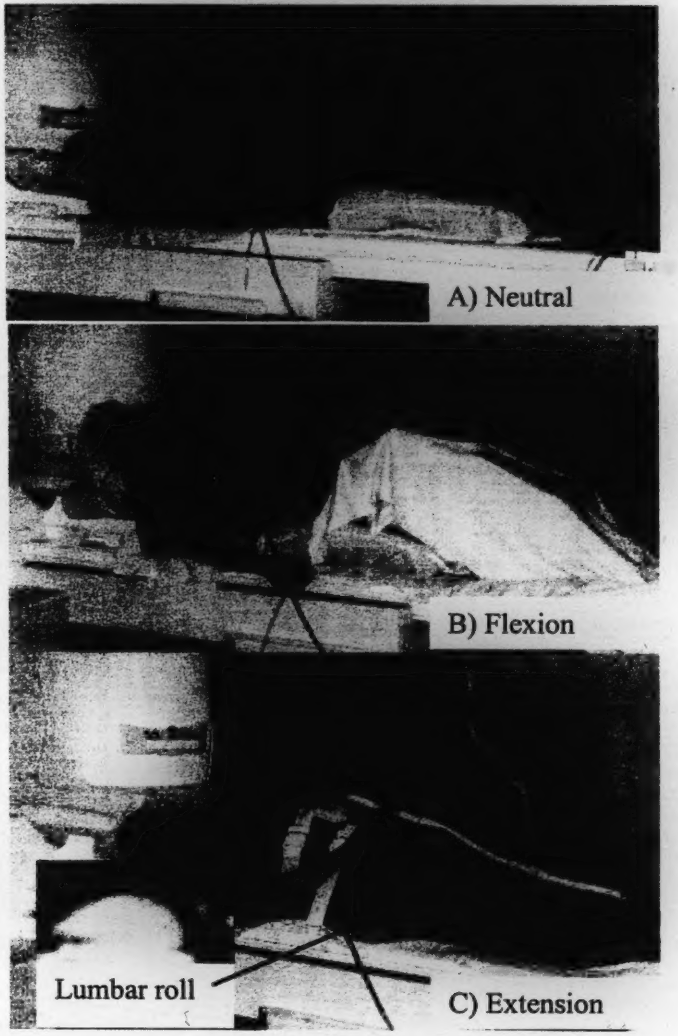
Disc	Flexion	Neutral	Extension	Overall p
	Anterior contour abnormality area (mm^2)			
L1L2	20.3 ± 10.3	21.7 ± 15.6	16.6 ± 8.4 ac	0.042
L2L3	30.7 ± 16.2	26.1 ± 16.9	20.2 ± 9.8 ac	< 0.001
L3L4	35.5 ± 15.4a	28.4 ± 13.1	27.4 ± 12.2 c	0.001
L4L5	35.0 ± 17.7	32.6 ± 18.0	24.5 ± 12.0 ac	0.003
L5S1	34.1 ± 16.2a	28.5 ± 14.5	23.9 ± 12.5 ac	< 0.001
	Anterior contour abnormality distance (mm)			
L1L2	2.8 ± 0.9a	2.2 ± 1.2	1.8 ± 0.7 ac	< 0.001
L2L3	3.4 ± 1.1a	2.2 ± 1.0	1.8 ± 0.6 ac	< 0.001
L3L4	3.4 ± 1.1a	2.3 ± 1.0	2.1 ± 0.7 c	< 0.001
L4L5	3.1 ± 1.2a	2.5 ± 1.1	1.7 ± 0.7 ac	< 0.001
L5S1	3.0 ± 1.1a	2.3 ± 1.0	1.6 ± 0.7 ac	< 0.001

* Values are mean ± standard deviation in degrees.

Abbreviations: a = significantly different from the neutral position,

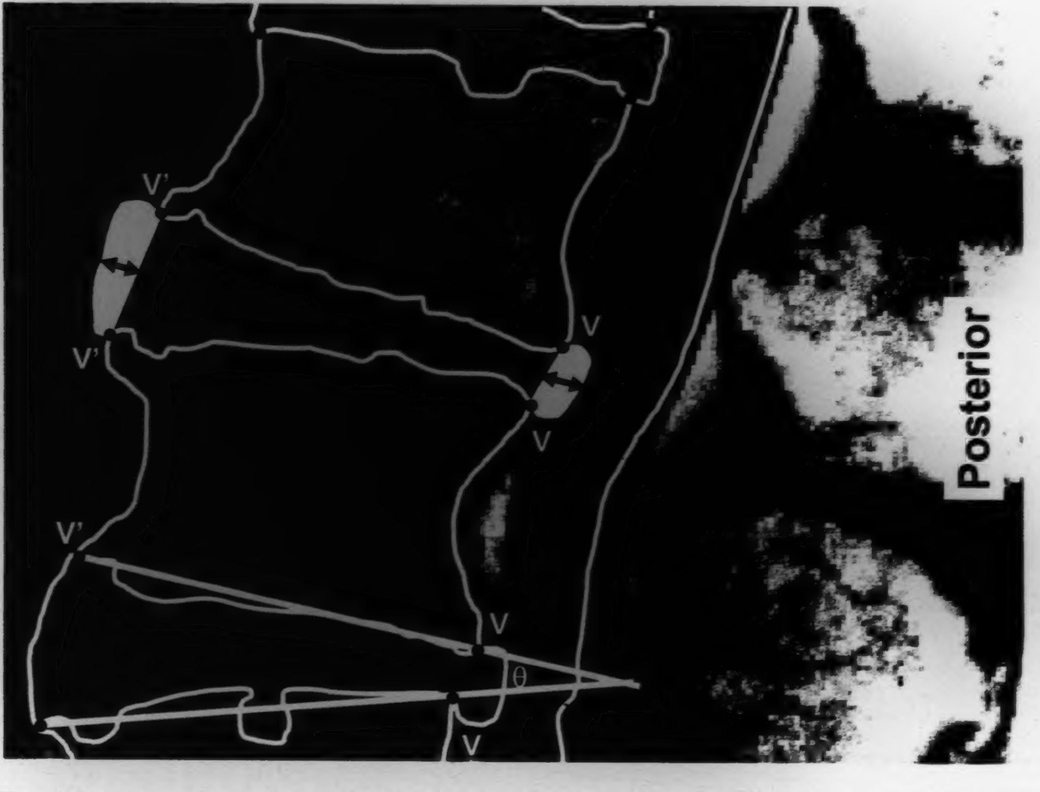
c = significant difference between the flexed and extended positions.($p < 0.05$)

Figure 4.1. Imaging position in neutral, flexion, and extension



Lumbar MR images (L1-S1) were obtained with the participant supine: a) in neutral position: neutral lumbar images were obtained with a positioning pillow under the knees, by using the spine array coils. b) in flexion: images with maximal lumbar flexion possible within the bore of the MR scanner were obtained with the shoulders, lower limbs, and pelvis supported with wedge positioning cushions, by using the spine array coils, and c) in extension: lumbar extension images were obtained by using a positioning roll placed under L3 with straps at the belt and ribs level to stabilize the participants and maintain a body coil in place.

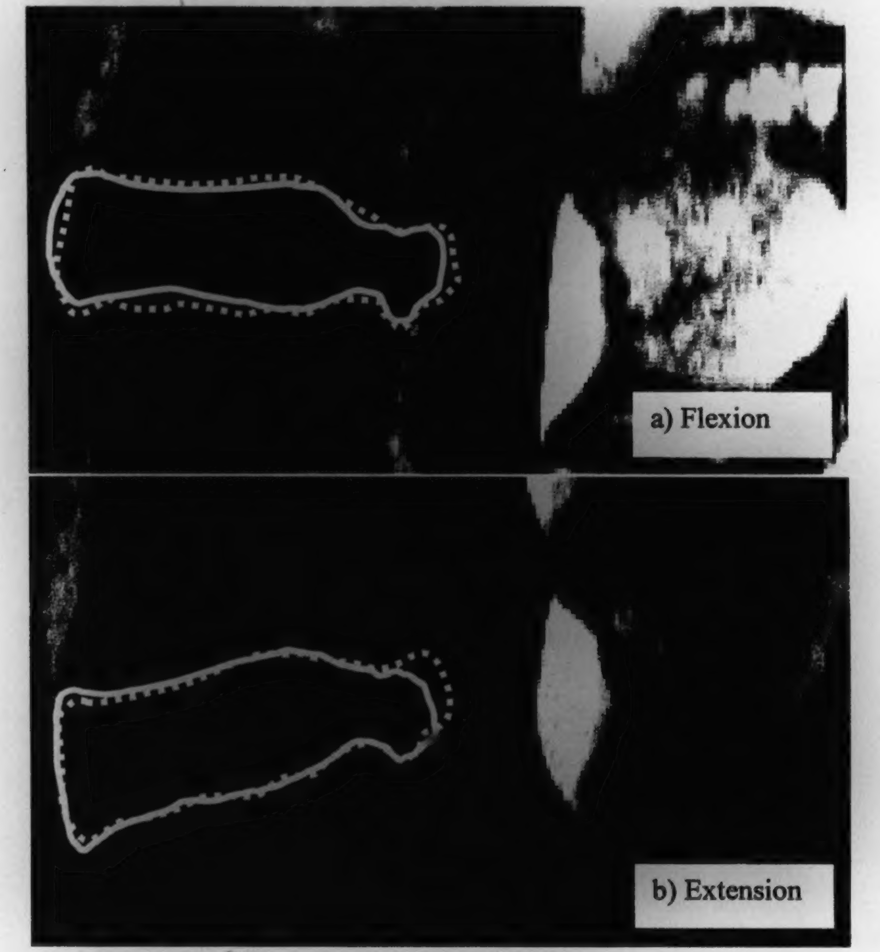
Figure 4.2. Measures of disc contour abnormality, using real and theoretical corners



Top: Measures of anterior and posterior area (White surface) and distance of disc contour abnormality, using vertebral corners as references (V) (double headed arrows).

Bottom: The "segmental angle" (θ) between the lines connecting the corners of the endplates above and below the disc.

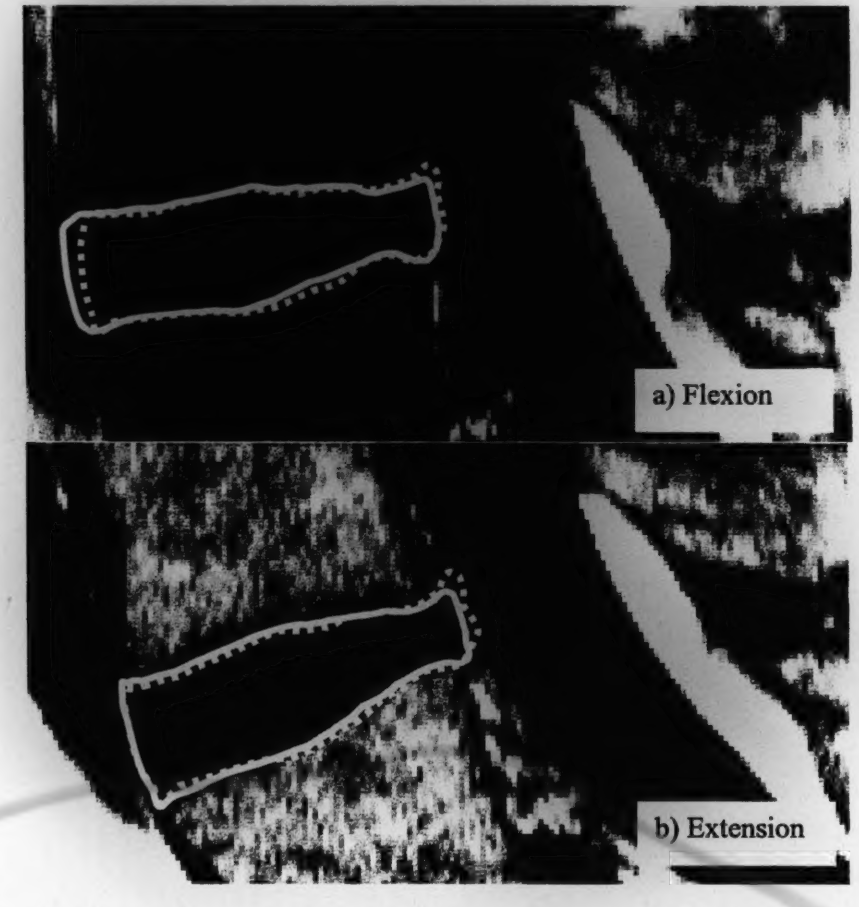
Figure 4.3. Example of a 57-year-old male with a large posterior contour abnormality in the neutral position and large decrease in quantitative disc contour abnormality in flexion or extension compared to neutral



a) L4L5 disc with posterior contour abnormality obtained with the spine in flexion is outlined with a continuous line. The outline copied from the image obtained in the neutral position is shown with a dotted line.

b) L4L5 disc with posterior contour abnormality obtained with the spine in extension is outlined with a continuous line. The outline copied from the image obtained in the neutral position is shown with a dotted line.

Figure 4.4. Example of a 35-year-old male with a small posterior contour abnormality in the neutral position and small decrease in quantitative disc contour abnormality in flexion or extension compared to the neutral position



a) L4L5 disc with minimum posterior contour abnormality obtained with the spine in flexion is outlined with a continuous line. The outline copied from the image obtained in the neutral position is shown with a dotted line.

b) L4L5 disc with minimum posterior contour abnormality obtained with the spine in extension is outlined with a continuous line. The outline copied from the image obtained in the neutral position is shown with a dotted line.

REFERENCES – CHAPTER 4

1. Mixter WJ, Barr JS. Rupture of the intervertebral disc with involvement of the spinal canal. *N Engl J Med.* 1934;1934(211):210-215.
2. Kuslich SD, Ulstrom CL, Michael CJ. The tissue origin of low back pain and sciatica: a report of pain response to tissue stimulation during operations on the lumbar spine using local anesthesia. *Orthop Clin North Am.* 1991;22(2):181-187.
3. Nakano M, Matsui H, Ishihara H et al. Serial changes of herniated intervertebral discs after posterior lumbar discectomy: the relation between magnetic resonance imaging of the postoperative intervertebral discs and clinical outcome. *J Spinal Disord.* 2001;14(4):293-300.
4. Mooney V. Presidential address. International Society for the Study of the Lumbar Spine. Dallas, 1986. Where is the pain coming from? *Spine.* 1987;12(8):754-759.
5. Frymoyer JW, Cats-Baril WL. An overview of the incidences and costs of low back pain. *Orthop Clin North Am.* 1991;22(2):263-271.
6. Merskey H, Bogduk N. *Classification for chronic pain syndromes and definitions of pain terms*. Second ed. Seattle, Washington: IASP Press, 1994,1-222.
7. Luoma K, Riihimaki H, Luukkonen R et al. Low back pain in relation to lumbar disc degeneration. *Spine.* 2000;25(4):487-492.
8. Smyth MJ, Wright V. Sciatica and the intervertebral disc. *J Bone Joint Surg Am.* 1958;40(6):1401-1418.
9. Fagerlund MK, Thelander U, Friberg S. Size of lumbar disc hernias measured using computed tomography and related to sciatic symptoms. *Acta Radiol.* 1990;31(6):555-558.
10. Komori H, Shinomiya K, Nakai O et al. The natural history of herniated nucleus pulposus with radiculopathy. *Spine.* 1996;21(2):225-229.
11. Olmarker K. Mechanical and biochemical injury of spinal nerve roots: an experimental perspective. In: Weinstein JN, Gordon SL, editors. *Low back pain A scientific and clinical overview*. Rosemont, Illinois: American Academy of Orthopaedic Surgeons, 1996: 215-233.
12. Modic MT, Masaryk TJ, Ross JS et al. Imaging of degenerative disk disease. *Radiology.* 1988;168(1):177-186.
13. Milette PC. Classification, diagnostic imaging, and imaging characterization of a lumbar herniated disk. *Radiol Clin North Am.* 2000;38(6):1267-1292.
14. Fardon DF, Milette PC. Nomenclature and classification of lumbar disc pathology. Recommendations of the Combined task Forces of the North American Spine Society, American Society of Spine Radiology, and American Society of Neuroradiology. *Spine.* 2001;26(5):E93-E113.
15. Brant-Zawadzki M, Jensen M. Spinal nomenclature. *Spine.* 1995;20(3):388-390.
16. Raininko R, Manninen H, Battie MC et al. Observer variability in the assessment of disc degeneration on magnetic resonance images of the lumbar and thoracic spine. *Spine.* 1995;20(9):1029-1035.
17. Battie MC, Videman T, Gibbons LE et al. 1995 Volvo Award in clinical sciences. Determinants of lumbar disc degeneration. A study relating lifetime exposures and

magnetic resonance imaging findings in identical twins. *Spine.* 1995;20(24):2601-2612.

18. Jarvik JG, Haynor DR, Koepsell TD et al. Interreader reliability for a new classification of lumbar disk disease. *Acad Radiol.* 1996;3(7):537-544.
19. Brant-Zawadzki MN, Jensen MC, Obuchowski N et al. Interobserver and intraobserver variability in interpretation of lumbar disc abnormalities. A comparison of two nomenclatures. *Spine.* 1995;20(11):1257-1263.
20. Milette PC, Fontaine S, Lepanto L et al. Differentiating lumbar disc protrusions, disc bulges, and discs with normal contour but abnormal signal intensity. Magnetic resonance imaging with discographic correlations. *Spine.* 1999;24(1):44-53.
21. Boos N, Rieder R, Schade V et al. 1995 Volvo Award in clinical sciences. The diagnostic accuracy of magnetic resonance imaging, work perception, and psychosocial factors in identifying symptomatic disc herniations. *Spine.* 1995;20(24):2613-2625.
22. Wood KB, Blair JM, Aepple DM et al. The natural history of asymptomatic thoracic disc herniations. *Spine.* 1997;22(5):525-529.
23. The influence of MRI image quality on the reliability of quantatitive measures of disc bulging/herniation. *5th Combined meeting of the Orthopaedic Research Societies of Canada, U.S.A., Japan and Europe.*; 04 Oct 10; 2004.310.
24. Jinkins JR, Dworkin J. Proceedings of the State-of-the-Art Symposium on Diagnostic and Interventional Radiology of the Spine, Antwerp, September 7, 2002 (Part two). Upright, weight-bearing, dynamic-kinetic MRI of the spine: pMRI/kMRI. *JBR -BTR.* 2003;86(5):286-293.
25. Vitzthum HE, Konig A, Seifert V. Dynamic examination of the lumbar spine by using vertical, open magnetic resonance imaging. *J Neurosurg.* 2000;93(1 Suppl):58-64.
26. Tanaka N, An HS, Lim TH et al. The relationship between disc degeneration and flexibility of the lumbar spine. *Spine J.* 2001;1(1):47-56.
27. Haughton VM, Rogers B, Meyerand ME et al. Measuring the axial rotation of lumbar vertebrae in vivo with MR imaging. *AJNR Am J Neuroradiol.* 2002;23(7):1110-1116.
28. Revel M, Mayoux-Benhamou MA, Aaron C et al. [Morphological variations of the lumbar foramina during flexion-extension and disk collapse]. *Rev Rhum Mal Osteoartic.* 1988;55(5):361-366.
29. Fujiwara A, An HS, Lim TH et al. Morphologic changes in the lumbar intervertebral foramen due to flexion-extension, lateral bending, and axial rotation: an in vitro anatomic and biomechanical study. *Spine.* 2001;26(8):876-882.
30. Schnebel BE, Watkins RG, Dillin W. The role of spinal flexion and extension in changing nerve root compression in disc herniations. *Spine.* 1989;14(8):835-837.
31. Reitman CA, Hipp JA, Kirking BC et al. Posterior annular strains during discography. *J Spinal Disord.* 2001;14(4):347-352.
32. Inufusa A, An HS, Lim TH et al. Anatomic changes of the spinal canal and intervertebral foramen associated with flexion-extension movement. *Spine.* 1996;21(21):2412-2420.

33. Botwin KP, Skene G, Tourres-Ramos FM et al. Role of weight-bearing flexion and extension myelography in evaluating the intervertebral disc. *Am J Phys Med Rehabil.* 2001;80(4):289-295.
34. Wildermuth S, Zanetti M, Duewell S et al. Lumbar spine: quantitative and qualitative assessment of positional (upright flexion and extension) MR imaging and myelography. *Radiology.* 1998;207(2):391-398.
35. Schnebel BE, Simmons JW, Chowning J et al. A digitizing technique for the study of movement of intradiscal dye in response to flexion and extension of the lumbar spine. *Spine.* 1988;13(3):309-312.
36. Lee J, Hida K, Seki T et al. An interspinous process distractor (X STOP) for lumbar spinal stenosis in elderly patients: preliminary experiences in 10 consecutive cases. *J Spinal Disord Tech.* 2004;17(1):72-77.
37. Beattie PF, Brooks WM, Rothstein JM et al. Effect of lordosis on the position of the nucleus pulposus in supine subjects. A study using magnetic resonance imaging. *Spine.* 1994;19(18):2096-2102.
38. Schmid MR, Stucki G, Duewell S et al. Changes in cross-sectional measurements of the spinal canal and intervertebral foramina as a function of body position: in vivo studies on an open-configuration MR system. *AJR Am J Roentgenol.* 1999;172(4):1095-1102.
39. Chung SS, Lee CS, Kim SH et al. Effect of low back posture on the morphology of the spinal canal. *Skeletal Radiol.* 2000;29(4):217-223.
40. Edmondston SJ, Song S, Bricknell RV et al. MRI evaluation of lumbar spine flexion and extension in asymptomatic individuals. *Man Ther.* 2000;5(3):158-164.
41. Weishaupt D, Schmid MR, Zanetti M et al. Positional MR imaging of the lumbar spine: does it demonstrate nerve root compromise not visible at conventional MR imaging? *Radiology.* 2000;215(1):247-253.
42. Fredericson M, Lee SU, Welsh J et al. Changes in posterior disc bulging and intervertebral foraminal size associated with flexion-extension movement: a comparison between L4-5 and L5-S1 levels in normal subjects. *Spine J.* 2001;1(1):10-17.
43. Zamani AA, Moriarty T, Hsu L et al. Functional MRI of the lumbar spine in erect position in a superconducting open-configuration MR system: preliminary results. *J Magn Reson Imaging.* 1998;8(6):1329-1333.
44. Fritz JM, Irrgang JJ. A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale. *Phys Ther.* 2001;81(2):776-788.
45. Videman T, Leskinen T, Stalhammar H et al. Effect of glycoaminoglycan polysulphate on the human intervertebral discs based on stature measurements. *Clin Biomech.* 1994;9(3):206-207.
46. Videman T, Nummi P, Battie MC et al. Digital assessment of MRI for lumbar disc desiccation. A comparison of digital versus subjective assessments and digital intensity profiles versus discogram and macroanatomic findings. *Spine.* 1994;19(2):192-198.
47. Parent E, Battie MC, Videman T. Quantitative MRI Measures of Disc Bulging/Herniation: Development and Reliability. *First Annual Bone and Joint Annual Meeting.* 1[1]. 10-24-2003.

48. Scientific Advisory Committee of the Medical Outcome Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res.* 2002;11(3):193-205.
49. Choy DS. Magnetic resonance imaging of the lumbosacral spine under compression. *J Clin Laser Med Surg.* 1997;15(2):71-73.
50. Stokes IA. Bulging of lumbar intervertebral discs: non-contacting measurements of anatomical specimens. *J Spinal Disord.* 1988;1(3):189-193.
51. Gallucci M, Bozzao A, Orlandi B et al. Does postcontrast MR enhancement in lumbar disk herniation have prognostic value? *J Comput Assist Tomogr.* 1995;19(1):34-38.
52. Ellenberg MR, Ross ML, Honet JC et al. Prospective evaluation of the course of disc herniations in patients with proven radiculopathy. *Arch Phys Med Rehabil.* 1993;74(1):3-8.
53. Kawaji Y, Uchiyama S, Yagi E. Three-dimensional evaluation of lumbar disc hernia and prediction of absorption by enhanced MRI. *J Orthop Sci.* 2001;6(6):498-502.
54. Ahn SH, Park HW, Byun WM et al. Comparison of clinical outcomes and natural morphologic changes between sequestered and large central extruded disc herniations. *Yonsei Med J.* 2002;43(3):283-290.
55. Ahn SH, Ahn MW, Byun WM. Effect of the transligamentous extension of lumbar disc herniations on their regression and the clinical outcome of sciatica. *Spine.* 2000;25(4):475-480.
56. Fujiwara A, Tamai K, An HS et al. The relationship between disc degeneration, facet joint osteoarthritis, and stability of the degenerative lumbar spine. *J Spinal Disord.* 2000;13(5):444-450.

CHAPTER 5: EVIDENCE OF CONVERGENT AND DIVERGENT VALIDITY FOR QUANTITATIVE MEASURES OF LUMBAR DISC CONTOUR ABNORMALITY ON MR IMAGES

INTRODUCTION

The intervertebral disc is the object of numerous diagnostic and therapeutic procedures in persons with low back pain (LBP). Disc contour abnormalities (bulging and herniation) are degenerative findings of the disc that have often been suspected of causing low back symptoms.[1-7] Medical specialists most frequently assess disc contour abnormalities by using qualitative ratings. However, quantitative measures have been shown to have superior reliability. Furthermore, quantitative measures have the potential to be particularly useful in follow-up and correlation studies because these measures have more measurement levels than qualitative rating scales, thereby increasing the likelihood of detecting change.[8-10]

Quantitative measures must also be valid. Of the few quantitative measures of disc contour abnormalities described in the literature to date,[3,11-25] some were assessed for reliability,[11,13,16-18,21,24,25] but few for validity.[14,19,24] Reliability has been discussed in the previous chapters. Quantitative measures do not distinguish between bulging and herniation and will be referred to as "quantitative measures of disc contour abnormality" in the remainder of this chapter, whose purpose is to discuss the validity of measures of contour abnormality and to provide validity evidence for quantitative measures of disc contour abnormality.

To my knowledge, validity studies of quantitative measures are rare and no criterion validity studies have been reported. Instead, construct validity studies of disc contour measurements such as correlations between measures obtained from different methods or different imaging planes have been conducted. Cooley et al. measured posterior disc contour abnormalities on printed sagittal image, selected by the rater as demonstrating the largest contour abnormality, by using a ruler and a backlit digitizer.

They found high inter-method correlations (Pearson's r = 0.97).[24] Holodny et al. measured quantitatively the volume of disc contour abnormality on axial and sagittal images and found an inter-method correlation of 0.85.[19] Thelander et al. compared two distance ratios, and one area ratio indicating the size of focal changes in disc morphology on CT images and found inter-method correlations between 0.73 and 0.97.[14] The correlations between changes over time measured by the different measures was also similar (0.74 to 0.82).[14]

The lack of evidence on the validity of quantitative measures may be attributable to many factors, the most important of which is probably the lack of an appropriate gold standard for criterion validity studies. Validity studies for qualitative measures are also rare.[26-36] The validity of qualitative ratings of the presence or absence of bulging or herniation has been assessed.[27-31,35,36] However, studies of the validity of magnitude ratings for abnormal disc contour are rare. In two series of 13 and 31 discs, MRI accuracy for the grade of disc contour abnormality assessed at surgery was 38% and 81%, respectively, and qualitative MRI assessment over- and under-estimated the grade of contour abnormality found at surgery.[29,37]

Criterion validity studies usually provide the strongest validity evidence, but selecting a criterion measure and conducting such studies would be difficult for quantitative measures of disc contour abnormality. Alternatively, validity evidence can be provided by using the convergent and divergent correlation approach. Convergent evidence of validity would be provided by strong correlations with different measures of the same construct and evidence of divergent validity by weak correlations with measures of a different construct. The qualitative assessment of MR images is the current standard of practice even though this assessment's limited reliability has been discussed in different studies.[9,24,38,39] Qualitative ratings of bulging and herniation can be obtained by using various definitions from different imaging planes.[39-45] The standardization of which images are read and which definitions are used is important. Comparing the correlation estimates with ratings from different medical specialists using different qualitative rating definitions would also increase the external generalizability of validity studies using qualitative ratings as a reference.

Quantitative measures of contour abnormality from sagittal images have been obtained mainly by using a single image[7,22-24,46,47] and often simply from the midsagittal image.[11,21,25] However, Ebeling and Reulen observed that 63.4% of focal disc contour abnormalities reported qualitatively were located medio-laterally, 19.8% laterally, and 11.4% within or lateral of the intervertebral compartment while only 5.4% affected the midline.[48] Axial images could be used to locate the lateral position of focal disc abnormalities and to select the sagittal slice where the largest changes occur. However, only one-third of focal contour abnormalities were found at the level of the disc space.[48] Frequently, disc material was found to be displaced caudally.[48] Thus, many contour abnormalities may be missed by relying solely on midsagittal or axial images. As a result, considering more than one imaging slice when measuring the size of disc contour abnormality quantitatively may be important. For this reason, alternatives to the use of only the mid-sagittal images will be studied. However, because obtaining quantitative measures from multiple slices is time-consuming, an evaluator should also determine whether all slices need to be considered to obtain a relevant estimate of quantitative disc contour abnormality or identify the subset of images most likely to show relevant contour abnormalities. The latter strategy would eliminate the need to rely on qualitative rater input when selecting which image to use to measure contour abnormality.

OBJECTIVES

1) To investigate convergent and divergent validity evidence to support the use of midsagittal quantitative measures of disc contour abnormality based on correlations between the quantitative values and medical specialists' qualitative ratings or amongst different quantitative measures of contour abnormality.

2) To compare the size of quantitative contour abnormality obtained from midsagittal or para-sagittal images and to determine on which slices quantitative contour abnormality should be measured.

HYPOTHESES

In the hypotheses below, the correlations expected to be moderate or high will provide convergent validity evidence, and the correlations expected to be low will provide evidence of divergent validity for the quantitative measures of disc contour abnormality. The rationale for each of the following hypothesis is provided in Table i), from Appendix F.

H1. Quantitative measures of posterior and anterior disc contour abnormality will have moderate correlations with qualitative measures of bulging, and low correlation with ratings of endplate irregularities.

H2. Quantitative measures of disc contour abnormality, using theoretical corners (as if no remodeling of the vertebrae due to degeneration had occurred), will correlate moderately, and measures using real corners will have relatively lower correlations with qualitative osteophyte ratings.

H3. Correlation coefficients between quantitative measures of contour abnormality and qualitative bulging for measures obtained from para-sagittal images will be equal or larger than correlation coefficients for measures obtained from midsagittal images. Correlations with qualitative measures of endplate irregularities and osteophytes will be similar when para-sagittal or midsagittal images are used.

H4. Areas and distances of quantitative posterior contour abnormality will correlate strongly together and correlations will be low with quantitative measures of anterior contour abnormality. The corresponding hypothesis also applies anteriorly.

H5. All quantitative measurement methods for contour abnormality (areas and distances using both types of reference corners) will correlate strongly together. However, similar measurement methods will correlate slightly more together (such as area with area, using real corners with using real corners) than with different methods.

H6. In subjects without moderate or severe qualitative ratings of bulging and without herniation, area and distance measures of disc contour abnormality will all correlate strongly together. However, in subjects identified by qualitative raters as having a herniation, which is typically asymmetric in the sagittal plane, areas will correlate more strongly together than with distances.

METHODS

Subjects

A first sample (Sample 1) of 132 men whose MR images had been assessed qualitatively by two or more raters was randomly selected from a population-based sample enrolled in 1992 and 1999 in a larger epidemiological study of twins.[49] A subset of 88 subjects from Sample 1 (Sample 1b) for which qualitative ratings were available from three raters was used for cross-validation analyses. A second sample (Sample 2) of 47 men was randomly selected from the same sample as Sample 1. Subjects with a higher severity of bulging and herniation were over-sampled to ensure adequate representation of a full range of severity. A third sample (Sample 3) of 285 men was also selected randomly from the same sample as Sample 1 and 2. Subjects were not selected on the basis of low back pain history (Samples 1, 2, and 3), or the presence or absence of clinical (Samples 1, 2, and 3), or imaging findings (Samples 1 and 3). This selection method ensured that the samples included a wide range of such findings. For each sample, the subjects' characteristics are presented in Table 5.1. All subjects received written information about the study and provided written consent before participating.

MR Imaging

MRI examinations of the lumbar spine (L1-S1) were conducted with participants lying supine, with knees supported, following 40 minutes in this position to reduce the influence of postural loading on disc morphology.[11,50-54] Before 1997, a Magnetom SP 4000 1.5T imager (Siemens AG, Germany) was used with a surface coil to obtain spin echo sagittal images of the lumbar spine (T12-S1) with a repetition time (TR) of 2600 ms and an echo time (TE) of 22 ms for proton-density and 90 ms for T2-weighted images (TR/TE = 2600/22-90). T1-weighted images were obtained by using a TR/TE of 650ms / 22ms. The slice thickness was 4 mm, and the inter-slice gaps were 0.4 mm. The acquisition matrix was 192 X 256 for T1 and 256 X 256 for PD/T2, with a field of view of 260 mm (pixel size 1.02 mm). Three excitations were averaged for T1, and only one excitation was obtained for PD/T2. The acquisition times for the T1-weigthed images and the dual echo PD/T2 images were 6:17 and 11:09 minutes, respectively. Axial PD and

T2-weighted images were also acquired by using 2450/22 and 2450/90 (TR/TE) sequences, respectively. Three axial images were available for the L2-L3 disc, and five were available for each disc included between L3 and S1. The slice thickness was 3 mm, and the inter-slice gaps were 0.3 mm. The acquisition matrix was 192 X 256, the field of view was 260 mm (pixel size 1.02 mm), and only one acquisition was obtained. The acquisition time for the dual echo axial image was 7:54 minutes. The phase encoding direction was anterior to posterior for both the axial and sagittal images.

After 1997, a Magnetom Vision 1.5T MR scanner (Siemens AG, Germany) was used to obtain sagittal T1, PD, and T2 images comparable to those of the Magnetom imager by using 650/12, 2600/20, and 2600/80 (TR/TE) sequences, respectively. The acquisition matrix was 256 X 512 for T1 and 168 X 256 for PD/T2. Two excitations were averaged for T1, and only one excitation was obtained for PD/T2. the acquisition times for the T1-weigthed images and the dual echo PD/T2 images were 5:36 and 7:20 minutes, respectively. Axial PD and T2-weighted images were also acquired by using 2200/20 and 2200/80 sequences, respectively. The acquisition time for the dual echo axial image was 7:06 minutes. The axial slice thickness was 4 mm, with a 0.4 mm gap. The other parameters remained the same.

Quantitative Assessment of MR Images

Custom-designed spine image analysis software (SpEx© v2.42) was used to obtain quantitative measures of disc contour abnormality. The contours of the anatomical boundaries of the lumbar discs, vertebrae, and spinal canal were segmented manually only on the midsagittal PD image. Fifteen to 20 minutes were required per subject. A single evaluator segmented images from Sample 1 and one of two evaluators segmented images for Sample 3 (Figure 5.1). The evaluator(s) selected the image where the spinous processes of L4 and L5 vertebrae were the most clearly revealed as the midsagittal image. For all subjects in Sample 2, all seven sagittal slices were segmented manually by the same evaluator. Approximately 105 to 140 minutes were required to segment each subject's images. The evaluators had prior experience manually segmenting over 200 spine MR images. The evaluators were blind to the participants' medical histories and to the qualitative ratings assigned to the images.

Segmentation Procedure

The evaluators vertically segmented the contour of the vertebral wall from the longitudinal ligament. To segment the contour of the disc annulus, at the endplate level the evaluators moved through the longitudinal ligament to follow the outer contour of the ligament/intervertebral disc. The contrast did not allow the annulus to be distinguished from the ligament, so the ligament was included in the disc area. To segment the disc from the vertebra, the evaluators followed horizontally the boundary between the vertebral bone and the endplates. Cartilaginous endplates were included in the disc since the contrast did not allow both tissues to be distinguished. The areas created by the intersection of those segmentation lines formed the regions of interest corresponding to the discs and vertebrae from which the measures were derived (Figure 5.1).[10,55-57] Segmented areas were then adjusted by using T2- (disc) and T1-weighted (vertebra) images and, if necessary, by taking advantage of different contrasts.

Quantitative Measures

The quantitative distances and areas of disc contour abnormality were then calculated by using the segmented outlines from two types of reference points (Figure 5.1). Disc contour abnormality was measured by using the "real" vertebral corners (Figure 5.1: R and R') as seen on the MR image at the intersection of the vertebral wall and the endplate contours. The second type of disc contour abnormality measures used the "theoretical" corners (Figure 5.1: T and T') of the vertebrae as a reference. "Theoretical corners" are defined by the intersections of a line connecting the diameter of the vertebra above the disc at 20% of the vertebra height to the corresponding diameter below the disc, with a line representing the disc-vertebra boundary (Figure 5.1). Measures using theoretical corners were developed to address the difficulty of precisely identifying the position of the real corners on MR images with low contrast and to measure disc contour abnormality when combined with the presence of osteophytes. The measures' reliability is extensively described in the previous chapters.[56,57]

Qualitative Assessment of MR images

Rationale We recognize that qualitative ratings do not represent a gold standard and these measures are not being used as such to study criterion validity. The identification of an appropriate criterion for the size of disc contour abnormality poses significant challenges. Furthermore, such a measurement could not be completed within this dissertation work because access to equipment and adequate cadaveric material was not possible within the planned duration of this thesis research program. There are also limitations to the use of cadaveric spines as a criterion standard as are discussed later. Instead, the qualitative ratings were used to provide evidence of construct validity. In the absence of a gold standard quantifying the extent of contour abnormality, medical specialists' qualitative assessment of MR images represents the best current practice and was selected as reference in the present validity study.

Raters and Qualitative Assessment Definitions Used for Samples 1 and 2

Independent raters evaluated the MR images qualitatively while blinded to the participants' clinical status and to the quantitative measurements. An experienced spine surgeon evaluated the images for all participants in Samples 1, 1b, and 2 by using a first set of rating definitions (Table 5.2), and evaluated all subjects in Sample 1b by using a second set of rating definitions (Table 5.3) for cross-validation. Also for cross-validation, a first experienced radiologist evaluated the images for all subjects in Sample 1b by using the second set of grading definitions (Table 5.3). A second experienced radiologist also evaluated the images for all subjects in Sample 1b by using a third set of grading definitions (Table 5.4).

Each intervertebral disc was evaluated separately based on standardized qualitative rating definitions for bulging (convergence with quantitative contour abnormality), endplate irregularities (divergence), and vertebral osteophytes (slightly higher correlation was expected with quantitative measures using theoretical corners) by using all sagittal images. The location of the largest focal disc bulges was determined by using the axial images.

Qualitative Data Entry

The author of this study entered all qualitative ratings from the spine surgeon and the general radiologist into a SPSS datasheet. Double entry was done for the first 58618 entries; the data entry error rate was 0.031%. Most errors were missing entries that were easily picked up during preliminary exploration of the data. As a result, a single entry was completed for the ratings from the spine surgeon and the first radiologist by using automatic detection of out-of-range entry. The data from the second radiologist had been entered for a previous project and were also submitted to detection of out-of-range entry.

Qualitative Ratings' Reliability

The intra-observer reliability of the clinicians' ratings used in the current study was reported previously for the images from a sub-sample of 20 twins re-evaluated within one month.[38] The rating definitions used in this reliability study are described in Table 5.3. Reliability varied depending on the spinal levels considered.[38,58] The inter-rater reliability estimates relevant for the current study are presented in Table 5.5.

Analyses

Descriptive Analyses

The means and standard deviations were calculated for each disc level for the new quantitative measures and subjects' continuous characteristics. The frequency distributions of the medical specialists' qualitative ratings are also reported.

Convergent and Divergent Validity Correlations

Campbell and Fiske[59], in 1959, proposed the multitrait-multimethod matrix approach to assess construct vality. Such a method is a useful alternative when a criterion validity study cannot be conducted. In order to claim that measures have construct validity using a multitrait-multimethod approach, one has to demonstrate both evidence of convergent and divergent validity. *Convergent validity* is the degree to which concepts that should be related theoretically are interrelated in reality.[60] In contrast, *divergent* *validity* is the degree to which concepts that should not be related theoretically are not interrelated in reality.[60]

Using this approach, a correlation matrix is obtained between all the measures used in the validation study. To conclude that the measures studied are valid, correlations between measures meant to asses the same construct/trait (convergent validity coefficients) should significantly differ from zero and be larger than correlations between measures meant to assess different construct (divergent validity coefficients).[61p155-156] Measures of the same construct using different methods should be high but measures of different constructs using the same methods should be low to show that what is measured is more important than how it is measured.[61p.156] Failure to obtain the pattern of correlation described above may be attributable to: (i) traits that should be distinct are not, (ii) methods are sensitive to common irrelevant variance, (iii) theory is incorrect or incomplete, (iv) the design is weak (e.g., small sample size), (v) the error of measurement is too high (leading to attenuation of the correlations), or (vi) a combination of the above.

Campbell and Fiske suggested that high convergent validity estimates were arguments to encourage further examination of the validity.[59] Recently, this validation approach has been used in rehabilitation and imaging studies and successfully contributed to the validation process of health questionnaires[62,63] and also of objective assessment procedures[60,64,65].

CONVERGENT AND DIVERGENT VALIDITY CORRELATIONS USING MIDSAGITTAL QUALITATIVE RATINGS

Spearman coefficients between quantitative measures and qualitative measures were calculated at each vertebral level separately. The data presented correspond to the average over the five lumbar levels. To obtain the average, correlation coefficients were first transformed into z-scores by using Fisher's procedure and then averaged.[66(p.362)] The mean z-score was then transformed back into the correlation coefficient by using the inverse Fisher transformation.[66(p.362)] This strategy prevents underestimation of the mean correlation coefficient due to the asymmetric distribution of the correlation coefficients.[66(p.362)]

For cross-validation (n = 88 from Sample 1b), the correlations were first computed by using the data from the repeated ratings by the orthopedic surgeon, who had used the definitions in Table 5.3. Second, the analyses were repeated by using the ratings from the first radiologist who had also used the definitions in Table 5.3. Third, the analyses were repeated by using the ratings from the other radiologist, who had used the definitions in Table 5.4. Cross-validation was deemed necessary because clinicians' qualitative ratings are known to have limited reliability,[38,57] which may lead to the attenuation of the correlation coefficients and limit the possibility of demonstrating convergent validity.[66(p.123-6)]

For the analyses above, in Sample 1b, qualitative ratings were obtained by the surgeon, who considered all the sagittal images although only the midsagittal images were analyzed for the quantitative measures. This discrepancy may have led to the attenuation of the correlation coefficients and was addressed by using two approaches. First, the correlation between the qualitative ratings of bulging and the quantitative measures of disc contour abnormality was studied only in a subset of subjects from Sample 1 for which the raters did not report bulging in other locations than the midsagittal. However, this method could be limited by a lower external generalizability or attenuation due to a restriction on the range (less contour abnormality included) because the prevalence of disc contour abnormalities on midsagittal images is reportedly smaller than that on para-sagittal images.[48] The second approach presented below also addressed this potential limitation.

CONVERGENT AND DIVERGENT VALIDITY CORRELATIONS USING MIDSAGITTAL AND PARA-SAGITTAL QUALITATIVE RATINGS

Quantitative measures were also obtained from all seven sagittal slices. Qualitative measures were recoded to obtain a qualitative rating for the midsagittal, left, and right findings by using the variable specifying the position of the degenerative findings from the axial images. The maximum quantitative measures on the three leftmost, rightmost, and most central sagittal slices were used as corresponding quantitative measures. Spearman correlations between the corresponding qualitative and quantitative measures were then obtained for the left, right, and central findings.

CONVERGENT AND DIVERGENT VALIDITY CORRELATIONS AMONGST THE QUANTITATIVE MEASURES OF DISC CONTOUR ABNORMALITY

The correlation between the quantitative measures of contour abnormality at each level was calculated by using Pearson's correlation coefficients. Correlations were first calculated by including all subjects in the sample, and then, by using sub-samples of subjects with and without bulging or herniation as identified by the orthopedic surgeon using the rating definitions from Table 5.2.

Differences in Quantitative Measures of Contour Abnormality between Sagittal Slices

The quantitative contour abnormality measurements on the para-sagittal slices were converted to percentages of the measurement obtained from the midsagittal slice, which was set at 100%. A repeated measures ANOVA was used with LSD post-hoc tests to study the differences in the disc contour abnormality dimensions among the five sagittal slices and the five lumbar disc levels. Differences were also tested in a subset of subjects with normal disc contour, and the results were compared to those from a subset of subjects with symmetrical discs (with bulging but no herniation).

RESULTS

Subjects' Description, Frequencies of Qualitative Ratings and Mean Quantitative Measures

Similar mean age, weight, height, and clinical severity indicators were observed for each sample and sub-sample (Table 5.1). With all rating definitions, the majority (> 86%) of the discs were rated as nearly normal (0 or 1) by all raters (Appendix G). By using the ratings in Table 5.3, more subjects were identified as "normal" compared to when the ratings in Table 5.2 or 5.4, even when a subject was rated by the orthopedic surgeon using the ratings from Table 5.2 and 5.3 (Appendix G). The means of the different quantitative measures of disc contour abnormality, using theoretical corners, were larger than the measures using real corners in all the samples considered (Appendix

H). The correlations reported hereafter were found to be lower when the mean qualitative ratings were not severe or when the variability was low (restriction of the range). The mean over the five lumbar disc levels and the standard deviation of the quantitative and qualitative measurement is included in the tables of correlation coefficients to facilitate interpretation.

Convergent-Divergent Validity Using Midsagittal Qualitative Ratings

Orthopedic Surgeon, Using Ratings from Table 5.2, Sample 1 (Table 5.6)

Quantitative measures correlated moderately (0.33-0.52) with corresponding anterior or posterior qualitative bulging ratings (H1). Anteriorly, the correlations of quantitative contour abnormality measures using theoretical corners were slightly higher than that of measures using real corners. Quantitative measures of contour abnormality did not correlate significantly with endplate irregularities (0.03-0.12) (H1). Anteriorly, measures using theoretical corners correlated more with osteophyte ratings (area 0.45, distance 0.54) than those using real corners (area 0.27, distance 0.31) (H2). Posteriorly, correlations between quantitative contour abnormality and qualitative osteophyte ratings were similar between the measures using the theoretical and real corners (theoretical area 0.19, distance 0.36; real area 0.17, distance 0.28) (H2).

Cross-Validation Using Data from Sample 1b

Orthopedic Surgeon, Using Ratings from Table 5.2 (Table 5.7)

Very similar convergent and divergent validity estimates were obtained from correlations using Sample 1 and the cross-validation Sample 1b (H1 and H2). This suggests that the different correlation estimates (reported hereafter) obtained when using data from different raters or different rating scales, are likely not due to the differences in sample size but, rather, to the different raters or rating scales used.

Orthopedic Surgeon, Using Definitions in Table 5.3: (Table 5.8)

The correlations obtained from Sample 1b when using the ratings obtained by the orthopedic surgeon when he used the definitions from Table 5.3 differed from the correlations obtained from Samples 1 and 1b by using the data obtained by the orthopedic surgeon when he used the definitions in Table 5.2 (Table 5.8 vs. Table 5.6 and 5.7). The quantitative measures correlated weakly (anterior 0.16-0.25; posterior 0.01-0.04) with the corresponding anterior or posterior qualitative bulging ratings (H1). The strength of the correlations with qualitative bulging was similar for the measures using either real or theoretical corners. The measures using theoretical corners correlated with the osteophyte ratings as weakly as the measures using real corners (Table 5.8) (H2) The weak correlations with the endplate irregularities were consistent with the correlations obtained in previous analyses (Tables 5.6 and 5.7) (H1).

First Radiologist, Using Ratings from Table 5.3 (Table 5.9)

The patterns of correlations obtained by using data from the radiologist who had also used the rating definitions in Table 5.3 were similar to the correlations obtained by the orthopedic surgeon when he used the rating definitions in Table 5.2 (Table 5.9 vs. Table 5.6 and 5.7). However, a comparison of the correlation estimates in Table 5.9 with Table 5.6 and 5.7 shows that, posteriorly, the quantitative contour abnormality measurements correlated more weakly with the qualitative bulging ratings (0.17-0.23) (H1). The correlations with the osteophyte ratings were also slightly weaker than those observed by the orthopedic surgeon when he used the rating definitions in Table 5.2, but were still stronger for the measures using theoretical corners than for the measures using real corners (H2).

Second Radiologist, Using Ratings from Table 5.4 (Table 5.10)

The second radiologist assessed only posterior bulges. The correlations between the qualitative ratings of bulging and the quantitative measures of contour abnormality by the second radiologist were also similar to those obtained by the orthopedic surgeon when he used the rating definitions in Table 5.2 (Table 5.10 compared to Table 5.6 and 5.7) (H1).

Convergent-Divergent Validity Using Midsagittal and Para-sagittal Qualitative Ratings

The correlations between the quantitative and qualitative measures, when considering the midsagittal (0.31-0.56) (Table 5.11), right side (0.12-0.42) (Table 5.12), and left side (0.17-0.41) (Table 5.13) findings separately, were not higher than the correlations described above for a single midsagittal image (0.33-0.52) (H3). The correlations between the endplate irregularities or osteophyte ratings with the quantitative contour abnormality measures from the para-sagittal images followed the same pattern as that described for the midsagittal image by using the rating definitions from Table 5.2 (H3).

Convergent-Divergent Validity Amongst Quantitative Measures

Subjects were classified as having a "normal" disc contour if the orthopedic surgeon, when using the definitions from Table 5.2, did not report a herniation rating larger than 0 or a bulging rating larger than 1 and were classified as having an "abnormal" disc contour otherwise. In the subset of subjects from Sample 3 with a normal disc contour (n = 212 to 259 depending on level, Table 5.14) as well as in a subset of subjects with abnormal disc contour (posteriorly n = 20 to 71, Table 5.15; anteriorly n = 24 to 63, depending on the disc level, Table 5.16), the correlations were similar in magnitude to the correlations that were obtained by using the whole sample (n = 285 in Sample 3, not shown).

Specifically, the correlations amongst the anterior or posterior quantitative bulging measures were above 0.75 (mean r = 0.86, range = 0.75-0.95, $p < 0.01$) (Table 5.14, 5.15, and 5.16) (H4). The correlations between the anterior and posterior measures were all smaller than 0.36 (mean r = 0.15, range = -0.03-0.36, $p < 0.01$) (H4). Anteriorly and posteriorly, the distance measures correlated slightly more strongly with distances, and areas correlated more strongly with areas (H5). The correlations between distance and area measures were slightly stronger for measures using the same reference corners (real with real and theoretical with theoretical) (H5). Posteriorly, the measures of disc contour abnormality correlated more strongly with the corresponding measures of area or distance than with the measures using the corresponding corners (H5). Although the differences were small, they consistently followed this pattern for posterior correlation. The pattern varied anteriorly.

For subjects with an abnormal disc contour, the strengths of the divergent correlations (between the anterior and posterior findings) were lower than in the analyses of the subset of subjects with a normal disc contour (Table 5.14) (mean 0.07, -0.03 to 0.22, $p > 0.01$, Tables 5.15, and 5.16) (H6). However, posteriorly and anteriorly, the mean convergent correlations over disc levels (among the anterior or posterior measures) did not change by a clinically important amount ($\Delta = -0.04$ to 0.05) (H6).

Differences in Quantitative Contour Abnormality Dimensions Between Sagittal Slices

The results related to this chapter's second objective are presented hereafter. In subjects with no moderate or severe bulging and no herniation determined qualitatively, midsagittal quantitative measurements of disc contour abnormality were significantly smaller than the measurements on all para-sagittal slices for all anterior measures. The midsagittal measures of posterior contour abnormality were smaller than those on at least three of four para-sagittal slices for all area measures and for the distance measures of posterior disc contour abnormality, using theoretical corners. The posterior contour abnormality measure on the rightmost para-sagittal slice was found to be smaller than the contour abnormality measure on the other para-sagittal slices when quantitified by the two area measures and smaller than only the leftmost slice when quantified by the distance measure using theoretical corners. No statistically significant interactions between the disc levels and the slice position, and no differences between lumbar levels were observed for any of the posterior disc contour abnormality measures calculated as percentages of the midsagittal measurement (Figure 5.2, left panel). For anterior disc contour abnormality, an interaction between the slice position and disc level was observed: the measures using the theoretical corners on the most para-sagittal slice at L5S1 were smaller than those at three of four of lumbar levels above. The midsagittal measures of anterior contour abnormality were also consistently smaller than those on the para-sagittal images. Depending on which quantitative measure was used, the quantitative anterior contour abnormality was found to be significantly larger on the most para-sagittal slice.

In subjects with symmetrical disc contours with moderate and large disc bulges a similar pattern of differences was observed. However, more statistically significant interactions between the slice position and disc levels were found for the anterior measures compared to the results from subjects with normal disc contour. Also, posteriorly, the measurements on the most para-sagittal slices were not as frequently found to be smaller than measurements on the slices adjacent to the midsagittal one (Figure 5.2, right panel).

DISCUSSION

Convergent and Divergent Validity with Midsagittal Qualitative Ratings.

Did the results provide validity evidence for quantitative measures? Overall, this study provided limited evidence of convergent and good evidence of divergent validity for quantitative measures of disc contour abnormality. The convergent and divergent hypotheses (H1 and H2) were formulated based on a study of which medical specialists' qualitative ratings clustered together.[67] These authors showed that ratings of disc bulging rarely clustered with ratings of endplate irregularities, and that osteophytes ratings clustered with ratings of disc bulging, depending on the disc levels considered.[67] Thus, the observation that the correlations were generally higher with corresponding qualitative measures of bulging than with endplate irregularity or osteophyte ratings supported the convergence hypothesis (H1). However, the strength of these convergent correlations was lower than expected.

The correlations with the qualitative ratings of endplate irregularities were low, supporting the divergent validity hypothesis (H1). Also, the correlations with the qualitative ratings of anterior osteophytes were slightly higher for measures using theoretical corners. However, the correlations with the osteophyte ratings were not as strong as hypothesized and, posteriorly, did not differ between the quantitative measures of contour abnormality, using real and theoretical corners (H2). Thus, the results only partially supported the hypothesis that osteophytes would be more strongly correlated with the measures using theoretical corners (H2).

Why were the correlations lower than expected? The strength of the correlational results may have been reduced by the low prevalence of severe qualitative ratings of bulging. Also, the majority of the subjects had minor or no endplate irregularities, and, at many lumbar levels, the raters did not find any posterior osteophytes. This low prevalence reduced the variance observed in the sample and, as a consequence, the magnitude of the correlation coefficients. The correlations may also have been attenuated because the qualitative ratings had only a limited number of increments while the quantitative measures were continuous scales and also because the qualitative ratings had relatively low reliability (Table 5.5) (Kappa and ICC: 0.40-0.76[38]). While correlation coefficients can be corrected for attenuation due to low reliability, correction has been shown to be inappropriate when the reliability is as low as that of the qualitative measures used in this study.

Limitations of qualitative ratings and strategies for corrections Averaging the qualitative ratings from multiple raters could theoretically have helped reduce the low reliability's influence on the correlations. However, analysis of the frequency distributions of the ratings by the three raters using the definitions from this study showed that the raters used the definitions differently, with the disagreements occurring mostly for the two lowest ratings of bulging. These differences suggested that the average of the the qualitative ratings would be difficult to interpret. The two lowest ratings may have been combined to allow for averaging the ratings from different raters and scales to prevent attenuation due to low reliability. However, this strategy would have also restricted the range of inter-subject variation in this sample. The analyses in which the two lowest ratings were either averaged or treated separately were compared to determine which approach helped avoid attenuation. The gain in reliability from the combination did not outweigh the attenuation due to the restriction on the range. The low prevalence of degeneration also limited the ability to determine and use the intra- and inter-rater reliabilities to correct the correlation coefficients for attenuation within the different samples because the lack of inter-subject variation also led to low reliabilities.

Consequently, repeating the analysis with each qualitative ratings dataset separately seemed the best approach for cross-validation.

The cross-validation results showed that the pattern of divergent correlation was consistent for all raters when correlating the quantitative contour abnormality measures with endplate irregularities and osteophytes (H1 and H2). Except for the data obtained from the first radiologist and from the orthopedic surgeon when they used the definition from Table 5.3, the data supported the hypotheses of convergent validity with bulging ratings (H1). This exception may have resulted because the raters using the definitions from Table 5.3 reported lower ratings for all degeneration findings using this set of definitions, leading to lower correlation coefficients due to the restriction on the range. Overall, the strength of the convergent validity estimates were lower than expected, but the results consistently were in the expected direction, so that they supported convergent validity hypotheses.

Convergent-Divergent Validity with Midsagittal and Para-sagittal Qualitative Ratings

Did the results provide validity evidence for quantitative measures? Overall, as for the validity evidence using only midsagittal images, the strength and pattern of the correlations between the quantitative contour abnormality measures and the qualitative ratings (i.e., bulging, endplate irregularities, and osteophytes) using measures from the para-sagittal images also provided limited evidence of convergent and good evidence of divergent validity. Furthermore, contrary to the hypothesis (H3) on the differences between correlations using data from the midsagittal or the para-sagittal images, the correlations with the qualitative ratings for para-sagittal bulging were lower rather than higher or equal to the correlations observed when only midsagittal images were used.

Why weren't the correlations with bulging ratings stronger in the para-sagittal planes? In addition to the attenuation due to the qualitative ratings' limited reliability, the correlations using data from the para-sagittal images may also have been attenuated by the low variance and low prevalence of the para-sagittal findings in the sample studied. Despite efforts to over-sample subjects with severe bulging/herniation findings, the prevalence of more severe bulging/herniations for images from the right and left side of the spine remained low.

Using the largest absolute quantitative measures of three slices may also have affected the correlations. Selecting the maximum from the three slices still may have led to taking the measure at another site than the one the qualitative raters used to assess the disc contour abnormality, thereby negatively affecting the correlations. Furthermore, the quantitative measures were not normalized to a slice-specific dimension such as total disc area or spinal canal dimension, which qualitative raters may consider when assigning disc bulging ratings.[68] The analysis of the differences between the slices in subjects with qualitatively determined normal disc contour suggests that normal disc morphology is larger in the para-sagittal planes. This observation suggests that qualitative raters may consider the disc contour as "normal" even when more bulging is present on a para-sagittal image. Such behaviour would also negatively affect the correlation between the qualitative and quantitative measures. Determining the clinical relevance of specific dimensions of contour abnormality measured from the midsagittal and para-sagittal slices is needed to decide whether considering larger morphological findings as normal para-sagittally is adequate.

Did subject selection influence the results of parallel validity studies? For the convergent and divergent validity studies using qualitative ratings, excluding subjects with herniation in other planes than midsagittal to study the correlations may have excluded an important portion of the sample. However, only a maximum of 1% (L1L2), 2% (L2L3), 3% (L3L4), 18% (L4L5), or 10% (L5S1) were excluded by using this strategy. Thus, it is unlikely to have resulted in the attenuation of the correlation coefficient by restricting the range of changes in disc morphology to a clinically important extent. Further, this strategy controlled for the potential attenuation of the correlation that could have arisen if findings on an image different from the one measured quantitatively had influenced the qualitative raters.

Convergent and Divergent Validity Amongst Quantitative Measures

Did the results provide validity evidence for quantitative measures? The results supported all the hypotheses regarding the correlations among quantitative measures (H4, H5 and H6). The strong correlations among the measures of posterior contour abnormality provided evidence of convergent validity (H4). The same conclusion applies to the measures of anterior contour abnormality (H4). The low correlations between measures of anterior and posterior contour abnormality provided evidence of divergent validity, supporting this study's hypotheses (H4). The strength of the correlations among the quantitative measures supports the hypothesis that the low correlations observed with qualitative measures are probably due to the smaller variance, smaller number of measure increments, and lower reliability of the qualitative measures.

Small differences ($\Delta < 0.17$) in the strength of the correlations between measures using the same and different types of corners support the hypothesis that measures using theoretical corners are measuring the same construct as measures using real corners: contour abnormality (H5). Furthermore, the observation that measures of distances or areas correlated only slightly more strongly with the other measures of distance or area rather than with the other measures using the same type of corners also supports the hypothesis that all the measures are addressing the same contour abnormality construct (H5). As a result, the decision of which measure to use should be determined by factors other than the current validity evidence, such as the reliability, ease of use, or further validation work.

What is the meaning of the differences between subjects with normal or abnormal disc contours? The results did not support the hypothesis that the correlation among area measures would increase in subjects with abnormal discs (H6). In subjects with abnormal disc contour, especially with herniation, the disc does not only expand posteriorly but also cephally and mostly caudally, as Ebeling and Reulen have reported.[48] In such instances, the area measure of contour abnormality would increase if the disc expanded asymmetrically in any of those directions. However, distance measures would detect an increase in contour abnormality only if the disc had expanded posteriorly. As a result, the correlations between the different area measures of disc contour abnormality would increase while the correlations between area and distances would be expected to decrease. These findings were not observed in the current study, possibly because only a small number of subjects had disc herniation and, possibly, because only a fraction of those had asymmetric findings.

Literature on the Validity of Quantitative Measures of Contour Abnormality

The measures of contour abnormality used in previous validity studies are different than the ones used in the current study.[14,19,24] Nevertheless, the strength of the inter-method correlations is comparable to the strength of those observed among different quantitative measures tested in the current study. The higher correlation observed by Cooley et al. may have resulted because their sample included subjects with more severe contour abnormalities, and because these researchers correlated the average of three measurements from each method rather than a single one, as was done in the current study.[24] No other study has reported a correlation between quantitative measures of contour abnormality and qualitative ratings.

Differences in Quantitative Contour Abnormality between Sagittal Slices

The findings from the current study suggest that the size of posterior contour abnormality is largest on the sagittal slices adjacent to the midsagittal images and that the size of the anterior contour abnormality is largest on the most para-sagittal slices. These findings were expected from the literature on the intervertebral disc's anatomy and pathophysiology.[48] In addition, a larger number of significant differences between the slices and disc levels were observed when the samples included subjects with moderate to severe disc bulging rather than only subjects with normal disc contour. This finding suggests that the quantitative measures of disc contour abnormality can detect the differences they are expected to measure. Unfortunately, the number of subjects with herniation findings was too small to analyze these subjects as a separate group to confirm that a larger number of differences between the slices and disc levels can be found in a group in which differences are expected.

Recommendations for Future Research

Validity of quantitative measures The current validity study used imperfect qualitative ratings, the current clinical standard, as a reference to provide validity evidence for quantitative measures of contour abnormality. Qualitative ratings cannot be considered a gold standard to be used in a criterion validity study. Thus, validity study designs to provide stronger validity evidence are discussed in this section. Such studies were not feasible as part of this dissertation. A proposal for a cadaver criterion validation study was prepared but the plan was later abandoned because an adequate sample of cadaver material could not be obtained within the normal duration of this PhD program. Furthermore, a cryomicrotome to obtain slices of the whole spine in the sagittal plane was not available. Co-investigators also had reservations about whether it would be possible to generalize results obtain from cadavers to in vivo research.

A criterion validity study could be conducted by using fresh human or animal cadaver spines. The technical challenges associated with such a study prevented the inclusion of fresh human or animal cadaver spines as part of this dissertation. First, several legal and practical challenges in Canada currently limit access to enough fresh human cadaver spines. Animal spines are easier to obtain, but animals are usually sacrificed at a young age and, as a result may present with a limited range in the severity of degenerative contour abnormalities. The second challenge is related to the measurement of contour abnormality in the spines collected. When using MRI, contour abnormality is measured from a virtual sagittal slice of the whole spine. Anatomical slices corresponding to those virtual slices are difficult to obtain, and the effect of freezing on the contour abnormality dimensions is unclear. The third challenge is the selection of a measurement method of contour abnormality on anatomical slices. Anatomical slices display the "true" anatomy as revealed on MR slices, and contour abnormality distance and area can be quantified in different ways. One option is to obtain a scaled high-resolution picture on which the contour abnormality distances and areas are quantified by using a ruler or digitization software to calculate the area. However, the latter involves manually segmenting the anatomical slice, and this procedure is prone to the same error as the method used to obtain measures from MR images. Alternatively, digitized pictures could be imported in the same custom spine analysis software, and measures could be obtained by using the same segmentation strategy. For the latter, staining anatomical slices may facilitate the segmentation process and reduce its associated error.

Alternatively, criterion validity could be assessed *in vivo* on subjects operated on because of disc contour abnormality. Measures of quantitative contour abnormality from MR images obtained before surgery could be compared with the quantification of contour abnormality during surgery or the quantification of the amount of tissue removed during surgery. However, perioperatively, the surgeon would observe the contour abnormality from a frontal plane, which is different from the sagittal plane from which the quantitative measures are obtained. Also, the amount of tissue removed during surgery does not necessarily correspond to the contour abnormality area measured on MR images, as surgeons sometimes remove epidural fat or additional disc tissue. Thus, the correlations between the surgical findings and the quantitative measures would not necessarily provide strong construct validity evidence.

The criterion validity of quantitative measurement procedures for disc contour abnormality could also be examined by using imaging phantoms created to have known dimensions of contour abnormality in the planes examined with MRI. Such a strategy would allow for the criterion validation of the measurement procedure if the phantoms could be made to display comparable contrast-to-noise levels as that on the images from the real subjects' lumbar spine. The challenge with such a strategy is to create phantoms whose images are similar to *in vivo* images.

Finally, an alternative validity study could compare the quantitative contour abnormality measures derived by using SpEx© to the measures obtained by using other segmentation procedures and measurement systems. However, to my knowledge, no other image analysis software currently derives measures of contour abnormality by using theoretical corners automatically from the outline of discs and vertebrae. Nevertheless, in such a validity study, numerous other software packages could allow the user to define a region of interest which could correspond to the areas defined using theoretical corners.

Use of quantitative measures from multiple slices in the study of determinants and signs and symptoms of disc contour abnormality It would be useful to recommend that not all slices are needed to obtain a clear estimation of quantitative disc contour abnormality because manual segmentation of multiple images is time-consuming. Until an automatic segmentation procedure becomes available for sagittal lumbar images, data-reduction strategies need to be considered. Because midsagittal images lead to smaller contour abnormality measurements than para-sagittal images (right and left), measuring only one para-sagittal image on each side of the spine may be an adequate compromise to save data-collection time and to gather the most relevant data. In research using measures from multiple slices, an investigator has to choose which measure of the magnitude of the contour abnormality to use in analyses: the largest, the average, or a sum from measurements on those images.

Another alternative to using a limited number of sagittal slices to obtain a contour abnormality measurement is to compute a contour abnormality volume measurement by using all sagittal slices, as suggested by Holodny et al. and Roberts et al.[19,54] The advantage would be that no information regarding the dimensions of the disc contour abnormality would be lost. The volume of the contour abnormality on one slice would be obtained by multiplying the corresponding area by the slice thickness plus the interslice gap.[14,19] The sum of changes in disc morphology volumes calculated for each slice could then be computed.[19] This alternative is more time-consuming than considering only the mid-sagittal image or a limited number of para-sagittal slices. Future research comparing the validity and reliability of different methods could assist in selecting the procedure to use for studying relationships between disc contour abnormality, and pain, or impairment measures.

CONCLUSIONS

Evidence of convergent and divergent validity partially supporting this study's hypotheses (H1, H2, and H3) was provided by using qualitative measures of various samples of subjects. Suggestions for future validation studies were described to address the limitations of using clinicians' qualitative ratings as a reference standard. Evidence of convergent and divergent validity strongly supporting the study's hypotheses (H4, H5) was provided by using quantitative measures of contour abnormality in various samples of subjects.

The quantitative measures of disc contour abnormality were found to be smallest in the midsagittal plane, and largest on the sagittal image adjacent to the midsagittal one posteriorly and on the most para-sagittal slice anteriorly. The differences in contour abnormality measurements between the sagital images and lumbar levels were found more frequently in subjects identified by qualitative raters as having herniation than in subjects with normal disc contours. Such findings suggest than quantitative measures are able to detect differences in disc contour abnormality when a difference is expected.

Further validation work addressing this study's limitations and challenges discussed in this chapter is needed to corroborate this study's findings and to provide more support for the value of the quantitative measures of contour abnormality.

Table 5.1. Subjects' description

Samples*	Age (yrs)	Height (cm)	Weight (kg)	Intensity of worst lifetime back pain (/100 worst)	Number of days of difficulty with work or leisure due to back problems
1 n=132	50 ± 8	175 ± 7	79 ± 10	59 ± 37	14 ± 57
1b n=88	49 ± 9	175 ± 6	80 ± 11	55 ± 37	20 ± 69
2 n=47	51 ± 7	174 ± 7	77 ± 12	59 ± 33	11 ± 54
3 n=285	49 ± 8	175 ± 7	80 ± 12	58 ± 36	12 ± 55

***Sample 1** was used for the correlation between the quantitative measures, and the qualitative ratings in the midsagittal plane by the orthopedist using rating definitions from Table 5.2.

Sample 1b was used in the cross-validation or the repetition of the correlation between the midsagittal qualitative and quantitative measures, with n=88 subjects from Sample 1 for which qualitative ratings were available from an orthopedist using rating definitions from Tables 5.2 and 5.3, a first radiologist using rating definitions from Table 5.3, and a second radiologist using definitions from Table 5.4.

Sample 2 was used for correlations between the qualitative measures by the orthopedist using rating definitions from Table 5.2, and the quantitative measures obtained from the midsagittal, right, and left sagittal images.

Sample 3 was used to study the correlations between the different quantitative measures of disc contour abnormality obtained from the midsagittal images.

Table 5.2. Qualitative rating scales used by the orthopedic surgeon in Sample 1, 1b, and 2

Vertebral upper and lower endplates irregularities
Schmorl's nodes
- 0: none present
- 1: slight defect (1-5 mm)
- 2: moderate defect (5-10 mm)
- 3: severe defect (>10 mm)

Disc bulging
- 0: none- normal contour of disc
- 1: slight - approx. 1.5 ± 1 mm
- 2: moderate - approx. 3.5 ± 1 mm
- 3: severe - $\geq$ 4.5 mm bulge

Vertebral osteophytes
- 0: none- normal contour
- 1: slight - approx. 1.5 ± 1 mm
- 2: moderate - approx. 3.5 ± 1 mm
- 3: severe - approx. $\geq$ 4.5 mm

Bulge / HNP – Axial view
- 0: none present
- 1: slight - approx. 1.5 ± 1 mm
- 2: moderate - approx. 3.5 ± 1 mm
- 3: severe - approx. 4.5 mm bulge or HNP
  - position noted by clock face position

Table 5.3. Qualitative rating scales used by the orthopedic surgeon and the first radiologist in Sample 1b and 2

Disc bulging

- 0: normal or equivocal
- 1: small: just notable: as large or larger than on the standard images
- 2: medium: as large or larger than on the standard images
- 3: large: as large or larger than on the standard images

If not symmetric, direction qualified with left, right, anterior, or posterior

Irregularities, nodes and defects in the vertebral endplates:

- 0: regular
- 1: small irregularities or one or two small nodes
- 2: multiple nodes or one large node (≥ 3 mm into the bone or in diameter)
- 3: larger damage, including changes other than Schmorl's nodes

Wavy but smooth endplates are regarded as regular.

Osteophytes in the lower and upper endplate (Rating and direction after maximal size)

- 0: regular or equivocal
- 1: notable, not large
- 2: large: as large or larger than on standard images
- 3: giant: as large or larger than on the standard images

Direction qualified with combinations of left or right, dorsal or ventral

Table 5.4. Qualitative rating scales used by the second radiologist in Sample 1b

Disc bulging

Y/N present or absent. If present:
1: mild <3mm
2: mod 3-5mm
3: severe > 5mm

Direction not qualified within this set of definitions. (rater focused on posterior changes)

Endplate disease: Not assessed by using this set of definitions.

Vertebral osteophytes: Not assessed by using this set of definitions.

Table 5.5. Reliability (ICC) of qualitative ratings in Sample 1b using definitions in Table 5.3

Qualitative rating	Inter-rater* L1-L4 (n=88)	Inter-rater L4-S1 (n=88)
Anterior bulging	0.44	0.21
Posterior bulging	0.02	0.27
Endplate irregularities	0.57	0.34
Osteophytes	0.65	0.58

* Orthopedist ratings using definitions from Table 5.2 vs first radiologist using ratings from Table 5.3.

Table 5.6. Mean correlation over all lumbar disc levels between quantitative measures of disc contour abnormality and qualitative ratings using definitions from Table 5.2 by an orthopedic surgeon in a sample of 132 subjects (Sample 1)

				Anterior bulging	Posterior bulging	Endplate irregularities	Vertebral osteophytes
			Means (SD)↓→	0.56 (0.69)	0.61 (0.67)	0.19 (0.55)	0.56 (0.82)
Anterior	Real corners	Area	109 (74)	0.33		0.03	0.27
		Distance	5 (2)	0.35		0.03	0.31
	Theoretical corners	Area	222 (101)	0.44		0.06	[illegible]
		Distance	9 (3)	0.51		0.06	[illegible]
Posterior	Real corners	Area	80 (49)		0.38	0.06	0.17
		Distance	5 (2)		0.47	0.07	0.28
	Theoretical corners	Area	98 (50)		0.41	0.08	0.19
		Distance	6 (2)		0.52	0.12	[illegible]

Note: Values are mean Spearman correlation coefficients over five lumbar disc levels.
Light grey shading illustrates convergence, moderate grey shading illustrates divergence, dark shading highlight tendency for higher correlations between measures using theoretical corners and osteophyte ratings.

9

Table 5.7. Mean correlation over all lumbar disc levels between quantitative measures of disc contour abnormality and qualitative ratings using definitions from Table 5.2 by an orthopedic surgeon in a sample of 88 subjects (Sample 1b)

				Anterior bulging	**Posterior bulging**	**Endplate irregularities**	**Vertebral osteophytes**
			Means (SD)↓→	0.57 (0.67)	0.57 (0.65)	0.18 (0.54)	0.51 (0.77)
Anterior	Real corners	Area	118 (69)	0.32		0.00	0.25
		Distance	5 (2)	0.33		0.00	0.28
	Theoretical corners	Area	220 (93)	0.43		0.01	0.40
		Distance	9 (3)	0.50		0.03	0.48
Posterior	Real corners	Area	84 (49)		0.41	0.07	0.25
		Distance	5 (2)		0.49	0.08	0.36
	Theoretical corners	Area	101 (51)		0.43	0.11	0.23
		Distance	6 (2)		0.52	0.14	0.37

Note: Values are mean Spearman correlation coefficients over five lumbar disc levels.
Light grey shading illustrates convergence, moderate grey shading illustrates divergence, dark shading highlight tendency for higher correlations between measures using theoretical corners and osteophyte ratings.

Table 5.8. Mean correlation over all lumbar disc levels between quantitative measures of disc contour abnormality and qualitative ratings using definitions from Table 5.3 by an orthopedic surgeon in a sample of 88 subjects (Sample 1b)

				Anterior bulging	**Posterior bulging**	**Endplate irregularities**	**Vertebral osteophytes**
			Means (SD)↓→	0.27 (0.68)	0.19 (0.44)	0.61 (0.93)	0.35 (0.79)
Anterior	Real corners	Area	118 (69)	0.16		0.10	0.20
		Distance	5 (2)	0.20		0.09	0.27
	Theoretical corners	Area	220 (93)	0.19		0.15	[illegible]
		Distance	9 (3)	0.25		0.20	[illegible]
Posterior	Real corners	Area	84 (49)		0.04	0.17	0.17
		Distance	5 (2)		0.01	0.21	0.21
	Theoretical corners	Area	101 (51)		0.04	0.18	[illegible]
		Distance	6 (2)		0.01	0.24	[illegible]

Note: Values are mean Spearman correlation coefficients over five lumbar disc levels.
Light grey shading illustrates convergence, moderate grey shading illustrates divergence, dark shading highlight tendency for higher correlations between measures using theoretical corners and osteophyte ratings.

Table 5.9. Mean correlation over all lumbar disc levels between quantitative measures of disc contour abnormality and qualitative ratings using definitions from Table 5.3 by the first radiologist in a sample of 88 subjects (Sample 1b)

				Anterior bulging	Posterior bulging	Endplate irregularities	Vertebral osteophytes
			Means (SD)↓→	0.39 (0.71)	0.17 (0.45)	0.40 (0.66)	0.30 (0.74)
Anterior	Real corners	Area	118 (69)	0.31		0.09	0.24
		Distance	5 (2)	0.29		0.08	0.24
	Theoretical corners	Area	220 (93)	0.39		0.11	0.31
		Distance	9 (3)	0.41		0.15	[illegible]
Posterior	Real corners	Area	84 (49)		0.20	0.15	0.15
		Distance	5 (2)		0.23	0.16	0.16
	Theoretical corners	Area	101 (51)		0.17	0.16	0.13
		Distance	6 (2)		0.22	0.20	0.18

Note: Values are mean Spearman correlation coefficients over five lumbar disc levels.
Light grey shading illustrates convergence, moderate grey shading illustrates divergence, dark shading highlight tendency for higher correlations between measures using theoretical corners and osteophyte ratings.

Table 5.10. Mean correlation over all lumbar disc levels between quantitative measures of disc contour abnormality and qualitative ratings using definitions from Table 5.4 by the second radiologist in a sample of 88 subjects (Sample 1b)

				Posterior bulging
			Means (SD)↓→	0.57 (0.73)
Anterior	Real corners	Area	118 (69)	0.19
		Distance	5 (2)	0.22
	Theoretical corners	Area	220 (93)	0.29
		Distance	9 (3)	0.37
Posterior	Real corners	Area	84 (49)	0.32
		Distance	5 (2)	0.45
	Theoretical corners	Area	101 (51)	0.34
		Distance	6 (2)	0.50

Note: Values are mean Spearman correlation coefficients over five lumbar disc levels.
Light grey shading illustrates convergence

Table 5.11. Mean correlation over all lumbar disc levels between the maximum quantitative measures of disc contour abnormality of the three most central sagittal images and qualitative ratings for the midsagittal images using definitions from Table 5.2 by the orthopedic surgeon in Sample 2 of 47 subjects

				Anterior midsagittal bulging	Posterior midsagittal bulging	Overall endplate irregularities	Overall vertebral osteophytes
			Means (SD)↓→	0.55 (0.71)	0.76 (0.60)	0.22 (0.49)	0.62 (0.79)
Anterior	Real corners	Area	132 (79)	0.31		0.10	0.38
		Distance	5 (3)	0.33		0.05	0.37
	Theoretical corners	Area	251 (96)	0.40		-0.01	[illegible]
		Distance	10 (3)	0.47		-0.02	[illegible]
Posterior	Real corners	Area	103 (45)		0.35	-0.07	0.28
		Distance	6 (2)		0.46	0.00	0.39
	Theoretical corners	Area	121 (49)		0.40	-0.02	[illegible]
		Distance	7 (2)		0.56	0.12	[illegible]

Note: Values are mean Spearman correlation coefficients over five lumbar disc levels.
Light grey shading illustrates convergence, moderate grey shading illustrates divergence, dark shading highlight tendency for higher correlations between measures using theoretical corners and osteophyte ratings.

Table 5.12. Mean correlation over all lumbar disc levels between the maximum quantitative measures of disc contour abnormality of the three sagittal images corresponding to the right side of the lumbar spine and qualitative ratings for the right side images using definitions from Table 5.2 by the orthopedic surgeon in Sample 2 of 47 subjects

				Anterior right bulging	Posterior right bulging	Overall endplate irregularities	Overall vertebral osteophytes
			Means (SD)↓→	0.76 (0.71)	0.76 (0.71)	0.22 (0.49)	0.62 (0.79)
Anterior	Real corners	Area	203 (123)	0.37		0.06	0.45
		Distance	7 (3)	0.35		0.05	0.44
	Theoretical corners	Area	352 (136)	0.36		0.03	0.49
		Distance	13 (4)	0.40		0.04	0.51
Posterior	Real corners	Area	129 (51)		0.12	0.05	0.20
		Distance	7 (2)		0.32	0.07	0.38
	Theoretical corners	Area	151 (54)		0.15	0.02	0.22
		Distance	9 (2)		0.42	0.04	0.45

Note: Values are mean Spearman correlation coefficients over five lumbar disc levels.
Light grey shading illustrates convergence, moderate grey shading illustrates divergence, dark shading highlight tendency for higher correlations between measures using theoretical corners and osteophyte ratings.

Table 5.13. Mean correlation over all lumbar disc levels between the maximum quantitative measures of disc contour abnormality of the three sagittal images corresponding to the left side of the lumbar spine and qualitative ratings for the left side images using definitions from Table 5.2 by the orthopedic surgeon in Sample 2 of 47 subjects

				Anterior left bulging	Posterior left bulging	Overall endplate irregularities	Overall vertebral osteophytes
			Means (SD)↓→	0.76 (0.71)	0.76 (0.71)	0.22 (0.49)	0.62 (0.79)
Anterior	Real corners	Area	223 (128)	0.26		[illegible]	0.41
		Distance	8 (3)	0.32		[illegible]	0.42
	Theoretical corners	Area	361 (150)	0.27		[illegible]	[illegible]
		Distance	13 (4)	0.40		[illegible]	[illegible]
Posterior	Real corners	Area	128 (52)		0.17	[illegible]	0.31
		Distance	7 (2)		0.32	[illegible]	0.37
	Theoretical corners	Area	150 (54)		0.21	[illegible]	[illegible]
		Distance	8 (2)		0.41	[illegible]	[illegible]

Note: Values are mean Spearman correlation coefficients over five lumbar disc levels.
Light grey shading illustrates convergence, moderate grey shading illustrates divergence, dark shading highlight tendency for higher correlations between measures using theoretical corners and osteophyte ratings.

Table 5.14. Mean Pearson correlation coefficients of five lumbar disc levels between quantitative measures of disc contour abnormality from the midsagittal image for subjects in Sample 3 with normal disc contour (no moderate or severe bulge and no herniation) (n=212 to 259 depending on disc level)

				ANTERIOR				POSTERIOR			
				Area		Distance		Area		Distance	
				Real	**Theo**	**Theo**	**Real**	**Real**	**Theo**	**Theo**	**Real**
			Means (SD)↓→	134 (64)	227 (84)	9 (3)	6 (2)	84 (40)	100 (43)	6 (2)	5 (2)
Anterior	Area	Real	134 (64)	1.00							
		Theo	227 (84)	0.90	1.00						
	Distance	Theo	9 (3)	0.78	0.88	1.00					
		Real	6 (2)	0.91	0.77	0.83	1.00				
Posterior	Area	Real	84 (40)	0.32	0.30	0.26	0.29	1.00			
		Theo	100 (43)	0.33	0.31	0.27	0.30	0.94	1.00		
	Distance	Theo	6 (2)	0.31	0.29	0.36	0.35	0.75	0.80	1.00	
		Real	5 (2)	0.31	0.28	0.33	0.34	0.84	0.76	0.90	1.00

Note: Values are mean Pearson's correlation coefficients over five lumbar disc levels.
Light grey shading illustrates convergence, moderate grey shading illustrates divergence

Table 5.15. Mean Pearson correlation coefficients of five lumbar disc levels between quantitative measures of disc contour abnormality from the midsagittal image for subjects in Sample 3 with abnormal posterior disc contour (moderate or severe bulge or herniation posteriorly) (n=20 to 71 depending on disc level)

				ANTERIOR				**POSTERIOR**			
				Area		**Distance**		**Area**		**Distance**	
				Real	**Theo**	**Theo**	**Real**	**Real**	**Theo**	**Theo**	**Real**
			Means (SD)↓→	203 (120)	328 (136)	13 (4)	8 (4)	131 (60)	149 (60)	10 (3)	9 (3)
Anterior	Area	Real	203 (120)	1.00							
		Theo	328 (136)	0.90	1.00						
	Distance	Theo	13 (4)	0.78	0.87	1.00					
		Real	8 (4)	0.93	0.80	0.86	1.00				
Posterior	Area	Real	131 (60)	0.01	0.03	0.09	0.06	1.00			
		Theo	149 (60)	0.02	0.05	0.09	0.06	0.97	1.00		
	Distance	Theo	10 (3)	0.06	0.06	0.22	0.18	0.75	0.78	1.00	
		Real	9 (3)	0.03	0.01	0.16	0.15	0.83	0.78	0.90	1.00

Note: Values are mean Pearson's correlation coefficients over five lumbar disc levels.
Light grey shading illustrates convergence, moderate grey shading illustrates divergence.

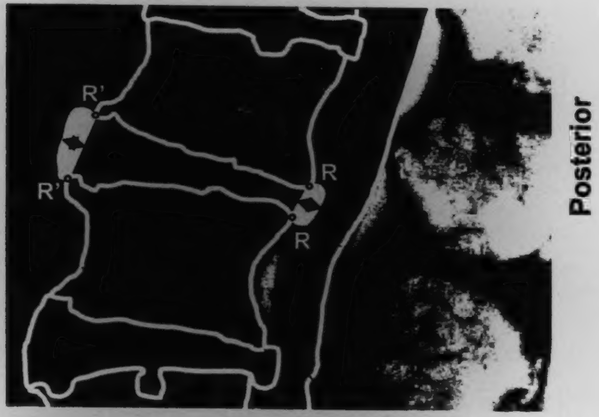
Table 5.16. Mean Pearson correlation coefficients of five lumbar disc levels between quantitative measures of disc contour abnormality from the midsagittal image for subjects in Sample 3 with abnormal anterior disc contour (moderate or severe bulge or herniation anteriorly) (n=24 to 63 depending on disc level)

				ANTERIOR				**POSTERIOR**			
				Area		**Distance**		**Area**		**Distance**	
				Real	**Theo**	**Theo**	**Real**	**Real**	**Theo**	**Theo**	**Real**
			Means (SD)↓→	201 (116)	303 (120)	12 (3)	8 (3)	120 (55)	141 (54)	10 (3)	8 (3)
Anterior	Area	Real	201 (116)	1.00							
		Theo	303 (120)	0.91	1.00						
	Distance	Theo	12 (3)	0.75	0.85	1.00					
		Real	8 (3)	0.92	0.80	0.85	1.00				
Posterior	Area	Real	120 (55)	-0.03	0.00	0.05	0.00	1.00			
		Theo	141 (54)	0.00	0.02	0.06	0.02	0.97	1.00		
	Distance	Theo	10 (3)	0.07	0.06	0.20	0.17	0.75	0.78	1.00	
		Real	8 (3)	0.01	-0.01	0.13	0.11	0.85	0.79	0.90	1.00

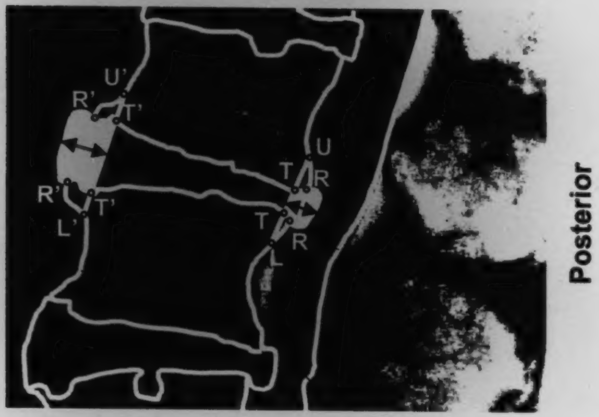
Note: Values are mean Pearson's correlation coefficients over five lumbar disc levels.
Light grey shading illustrates convergence, moderate grey shading illustrates divergence.

Figure 5.1. Measures of disc contour abnormality, using real and theoretical corners

Using Real Corners



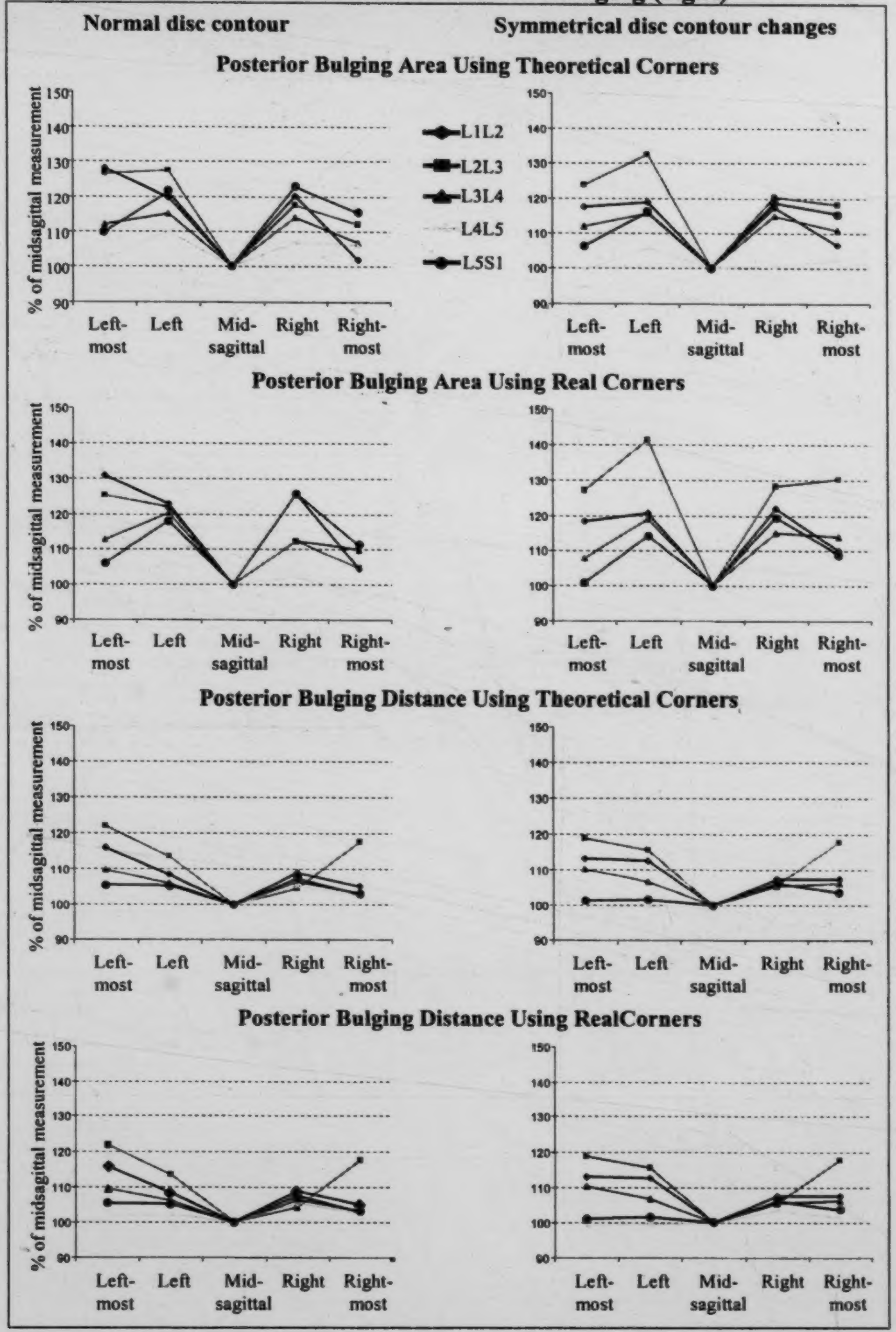
Using Theoretical Corners



Top: Measures of anterior and posterior area of disc contour abnormality, using real corners (R) (White surface) and measures of anterior and posterior distance of disc contour abnormality using real corners as references (R) (double headed arrows).

Bottom: Measures of anterior and posterior area of disc contour abnormality, using theoretical corners (T) (White surface) and measures of anterior and posterior distance of disc contour abnormality using theoretical corners as references (T) (double-headed arrows). Theoretical corners are defined by the intersections (T) of a line connecting the diameter of the vertebra above the disc at 20% of the vertebra height (U) to the corresponding diameter below the disc (L) with a line representing the disc-vertebra boundary.

Figure 5.2. Profile of contour abnormality dimensions for each level depending on which sagittal image is used to obtain measurement. Subject from Sample 2 qualitatively identified as having normal disc contour (left) compared to subjects known to have some levels with moderate or severe bulging (right)



REFERENCES – CHAPTER 5

1. Mixter WJ, Barr JS. Rupture of the intervertebral disc with involvement of the spinal canal. *N Engl J Med.* 1934;1934(211):210-215.
2. Kuslich SD, Ulstrom CL, Michael CJ. The tissue origin of low back pain and sciatica: a report of pain response to tissue stimulation during operations on the lumbar spine using local anesthesia. *Orthop Clin North Am.* 1991;22(2):181-187.
3. Nakano M, Matsui H, Ishihara H et al. Serial changes of herniated intervertebral discs after posterior lumbar discectomy: the relation between magnetic resonance imaging of the postoperative intervertebral discs and clinical outcome. *J Spinal Disord.* 2001;14(4):293-300.
4. Mooney V. Presidential address. International Society for the Study of the Lumbar Spine. Dallas, 1986. Where is the pain coming from? *Spine.* 1987;12(8):754-759.
5. Frymoyer JW, Cats-Baril WL. An overview of the incidences and costs of low back pain. *Orthop Clin North Am.* 1991;22(2):263-271.
6. Merskey H, Bogduk N. *Classification for chronic pain syndromes and definitions of pain terms*. Second ed. Seattle, Washington: IASP Press, 1994,1-222.
7. Luoma K, Riihimaki H, Luukkonen R et al. Low back pain in relation to lumbar disc degeneration. *Spine.* 2000;25(4):487-492.
8. Bird P, Ejbjerg B, McQueen F et al. OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Studies. Exercise 5: an international multicenter reliability study using computerized MRI erosion volume measurements. *J Rheumatol.* 2003;30(6):1380-1384.
9. Parent E, Videman T, Rogers T et al. The influence of MRI image quality on the reliability of quantitative measures of disc bulging/herniation. *5th Combined Meeting of the Orthopaedic Research Societies of Canada, U.S.A., Japan and Europe.* 310. 10-10-2004.
10. Battie MC, Videman T, Gibbons LE et al. 1995 Volvo Award in clinical sciences. Determinants of lumbar disc degeneration. A study relating lifetime exposures and magnetic resonance imaging findings in identical twins. *Spine.* 1995;20(24):2601-2612.
11. Fredericson M, Lee SU, Welsh J et al. Changes in posterior disc bulging and intervertebral foraminal size associated with flexion-extension movement: a comparison between L4-5 and L5-S1 levels in normal subjects. *Spine J.* 2001;1(1):10-17.
12. Henmi T, Sairyo K, Nakano S et al. Natural history of extruded lumbar intervertebral disc herniation. *J Med Invest.* 2002;49(1-2):40-43.
13. Carragee EJ, Kim DH. A prospective analysis of magnetic resonance imaging findings in patients with sciatica and lumbar disc herniation. Correlation of outcomes with disc fragment and canal morphology. *Spine.* 1997;22(14):1650-1660.
14. Thelander U, Fagerlund M, Friberg S et al. Describing the size of lumbar disc herniations using computed tomography. A comparison of different size index calculations and their relation to sciatica. *Spine.* 1994;19(17):1979-1984.

15. Pneumaticos SG, Hipp JA, Esses SI. Sensitivity and specificity of dural sac and herniated disc dimensions in patients with low back-related leg pain. *J Magn Reson Imaging.* 2000;12(3):439-443.
16. Tonami H, Yokota H, Nakagawa T et al. Percutaneous laser discectomy: MR findings within the first 24 hours after treatment and their relationship to clinical outcome. *Clin Radiol.* 1997;52(12):938-944.
17. Wood KB, Blair JM, Aepple DM et al. The natural history of asymptomatic thoracic disc herniations. *Spine.* 1997;22(5):525-529.
18. Buttermann GR. Lumbar disc herniation regression after successful epidural steroid injection. *J Spinal Disord Tech.* 2002;15(6):469-476.
19. Holodny AI, Kisza PS, Contractor S et al. Does a herniated nucleus pulposus contribute significantly to a decrease in height of the intervertebral disc? Quantitative volumetric MRI. *Neuroradiology.* 2000;42(6):451-454.
20. Bozzao A, Gallucci M, Masciocchi C et al. Lumbar disk herniation: MR imaging assessment of natural history in patients treated without surgery. *Radiology.* 1992;185(1):135-141.
21. Harada A, Okuizumi H, Miyagi N et al. Correlation between bone mineral density and intervertebral disc degeneration. *Spine.* 1998;23(8):857-861.
22. Sahlstrand T, Lonntoft M. A prospective study of preoperative and postoperative sequential magnetic resonance imaging and early clinical outcome in automated percutaneous lumbar discectomy. *J Spinal Disord.* 1999;12(5):368-374.
23. Matsubara Y, Kato F, Mimatsu K et al. Serial changes on MRI in lumbar disc herniations treated conservatively. *Neuroradiology.* 1995;37(5):378-383.
24. Cooley JR, Danielson CD, Schultz GD et al. Posterior disk displacement: morphologic assessment and measurement reliability-lumbar spine. *J Manipulative Physiol Ther.* 2001;24(5):317-326.
25. Luoma K, Riihimaki H, Raininko R et al. Lumbar disc degeneration in relation to occupation. *Scand J Work Environ Health.* 1998;24(5):358-366.
26. Pfirrmann CW, Dora C, Schmid MR et al. MR image-based grading of lumbar nerve root compromise due to disk herniation: reliability study with surgical correlation. *Radiology.* 2004;230(2):583-588.
27. Maravilla KR, Lesh P, Weinreb JC et al. Magnetic resonance imaging of the lumbar spine with CT correlation. *AJNR Am J Neuroradiol.* 1985;6(2):237-245.
28. Forristall RM, Marsh HO, Pay NT. Magnetic resonance imaging and contrast CT of the lumbar spine. Comparison of diagnostic methods and correlation with surgical findings. *Spine.* 1988;13(9):1049-1054.
29. Greenspan A, Amparo EG, Gorczyca DP et al. Is there a role for diskography in the era of magnetic resonance imaging? Prospective correlation and quantitative analysis of computed tomography-diskography, magnetic resonance imaging, and surgical findings. *J Spinal Disord.* 1992;5(1):26-31.
30. Birney TJ, White JJ, Jr., Berens D et al. Comparison of MRI and discography in the diagnosis of lumbar degenerative disc disease. *J Spinal Disord.* 1992;5(4):417-423.
31. Janssen ME, Bertrand SL, Joe C et al. Lumbar herniated disk disease: comparison of MRI, myelography, and post-myelographic CT scan with surgical findings. *Orthopedics.* 1994;17(2):121-127.

32. Platenberg RC, Hubbard GB, Ehler WJ et al. Spontaneous disc degeneration in the baboon model: magnetic resonance imaging and histopathologic correlation. *J Med Primatol.* 2001;30(5):268-272.
33. Huang TS, Zucherman JF, Hsu KY et al. Gadopentetate dimeglumine as an intradiscal contrast agent. *Spine.* 2002;27(8):839-843.
34. Wittenberg RH, Lutke A, Longwitz D et al. The correlation between magnetic resonance imaging and the operative and clinical findings after lumbar microdiscectomy. *Int Orthop.* 1998;22(4):241-244.
35. Schulz KS, Walker M, Moon M et al. Correlation of clinical, radiographic, and surgical localization of intervertebral disc extrusion in small-breed dogs: a prospective study of 50 cases. *Vet Surg.* 1998;27(2):105-111.
36. Silverman CS, Lenchik L, Shimkin PM et al. The value of MR in differentiating subligamentous from supraligamentous lumbar disk herniations. *AJNR Am J Neuroradiol.* 1995;16(3):571-579.
37. Kim KY, Kim YT, Lee CS et al. MRI classification of lumbar herniated intervertebral disc. *Orthopedics.* 1992;15(4):493-497.
38. Raininko R, Manninen H, Battie MC et al. Observer variability in the assessment of disc degeneration on magnetic resonance images of the lumbar and thoracic spine. *Spine.* 1995;20(9):1029-1035.
39. Brant-Zawadzki MN, Jensen MC, Obuchowski N et al. Interobserver and intraobserver variability in interpretation of lumbar disc abnormalities. A comparison of two nomenclatures. *Spine.* 1995;20(11):1257-1263.
40. Parkkola R, Rytokoski U, Kormano M. Magnetic resonance imaging of the discs and trunk muscles in patients with chronic low back pain and healthy control subjects. *Spine.* 1993;18(7):830-836.
41. Jarvik JG, Haynor DR, Koepsell TD et al. Interreader reliability for a new classification of lumbar disk disease. *Acad Radiol.* 1996;3(7):537-544.
42. Boos N, Semmer N, Elfering A et al. Natural history of individuals with asymptomatic disc abnormalities in magnetic resonance imaging: predictors of low back pain-related medical consultation and work incapacity. *Spine.* 2000;25(12):1484-1492.
43. Karppinen J, Malmivaara A, Tervonen O et al. Severity of symptoms and signs in relation to magnetic resonance imaging findings among sciatic patients. *Spine.* 2001;26(7):E149-E154.
44. Milette PC, Fontaine S, Lepanto L et al. Differentiating lumbar disc protrusions, disc bulges, and discs with normal contour but abnormal signal intensity. Magnetic resonance imaging with discographic correlations. *Spine.* 1999;24(1):44-53.
45. Jarvik JJ, Hollingworth W, Heagerty P et al. The longitudinal assessment of imaging and disability of the back (laidback) study: baseline data. *Spine.* 2001;26(10):1158-1166.
46. Silcox DH, III, Horton WC, Silverstein AM. MRI of lumbar intervertebral discs. Diurnal variations in signal intensities. *Spine.* 1995;20(7):807-811.
47. Kato F, Mimatsu K, Kawakami N et al. Serial changes observed by magnetic resonance imaging in the intervertebral disc after chemonucleolysis. A consideration of the mechanism of chemonucleolysis. *Spine.* 1992;17(8):934-939.

48. Ebeling U, Reulen HJ. Are there typical localisations of lumbar disc herniations? A prospective study. *Acta Neurochir (Wien).* 1992;117(3-4):143-148.
49. Simonen R, Levalahti E, Kaprio J et al. Multivariate Genetic Analysis of Lifetime Exercise and Environmental Factors. *Med Sci Sports Exerc.* 2004;36(9):1559-1566.
50. Botsford DJ, Esses SI, Ogilvie-Harris DJ. In vivo diurnal variation in intervertebral disc volume and morphology. *Spine.* 1994;19(8):935-940.
51. Malko JA, Hutton WC, Giacometti AR et al. Do diurnal changes in loading affect the interpretation of MRI scans of the lumbar spine? *J Spinal Disord.* 1996;9(2):129-135.
52. Malko JA, Hutton WC, Fajman WA. An in vivo magnetic resonance imaging study of changes in the volume (and fluid content) of the lumbar intervertebral discs during a simulated diurnal load cycle. *Spine.* 1999;24(10):1015-1022.
53. Park CO. Diurnal variation in lumbar MRI. Correlation between signal intensity, disc height, and disc bulge. *Yonsei Med J.* 1997;38(1):8-18.
54. Roberts N, Hogg D, Whitehouse GH et al. Quantitative analysis of diurnal variation in volume and water content of lumbar intervertebral discs. *Clin Anat.* 1998;11(1):1-8.
55. Videman T, Nummi P, Battie MC et al. Digital assessment of MRI for lumbar disc desiccation. A comparison of digital versus subjective assessments and digital intensity profiles versus discogram and macroanatomic findings. *Spine.* 1994;19(2):192-198.
56. Parent E, Battie MC, Videman T. Quantitative MRI Measures of Disc Bulging/Herniation: Development and Reliability. *First Annual Bone and Joint Annual Meeting.* 1[1]. 10-24-2003.
57. The influence of MRI image quality on the reliability of quantatitive measures of disc bulging/herniation. *5th Combined meeting of the Orthopaedic Research Societies of Canada, U.S.A., Japan and Europe.*; 04 Oct 10; 2004.310.
58. Battie MC, Haynor DR, Fisher LD et al. Similarities in degenerative findings on magnetic resonance images of the lumbar spines of identical twins. *J Bone Joint Surg Am.* 1995;77(11):1662-1670.
59. Campbell DT, Fiske DW. Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. *Psychological Bulletin.* 1959;56(2):81-105.
60. Frisoni GB, Geroldi C, Beltramello A et al. Radial width of the temporal horn: a sensitive measure in Alzheimer disease. *AJNR Am J Neuroradiol.* 2002;23(1):35-47.
61. Streiner DL, Norman GR. *Health measurement scales: A practical guide to their development and use.* Second ed. New York, New York: Oxford Medical Publications, 1995,1-231.
62. Martin BC, Pathak DS, Sharfman MI et al. Validity and reliability of the migraine-specific quality of life questionnaire (MSQ Version 2.1). *Headache.* 2000;40(3):204-215.
63. Pransky G, Feuerstein M, Himmelstein J et al. Measuring functional outcomes in work-related upper extremity disorders. Development and validation of the Upper Extremity Function Scale. *J Occup Environ Med.* 1997;39(12):1195-1202.

64. Lin MR, Hwang HF, Hu MH et al. Psychometric comparisons of the timed up and go, one-leg stand, functional reach, and Tinetti balance measures in community-dwelling older people. *J Am Geriatr Soc.* 2004;52(8):1343-1348.
65. Geroldi C, Galluzzi S, Testa C et al. Validation study of a CT-based weighted rating scale for subcortical ischemic vascular disease in patients with mild cognitive deterioration. *Eur Neurol.* 2003;49(4):193-209.
66. Glass GV, Hopkins KD. *Statistical methods in education and psychology.* 3 ed. Needham Heights, MA, USA: Allyn and Bacon, 1995,1-674.
67. Videman T, Battie MC, Gill K et al. Magnetic resonance imaging findings and their relationships in the thoracic and lumbar spine. Insights into the etiopathogenesis of spinal degeneration. *Spine.* 1995;20(8):928-935.
68. Davatzikos C, Liu D, Shen D et al. Spatial normalization of spine MR images for statistical correlation of lesions with clinical symptoms. *Radiology.* 2002;224(3):919-926.

CHAPTER 6: GENERAL DISCUSSION

The main objective of this thesis was to develop quantitative measures of disc contour abnormality on MR images, assess their measurement properties, and compare them to qualitative measures of bulging and herniation. Four different types of measures of anterior and posterior lumbar disc contour abnormalities were proposed and evaluated. Area and distance measures of disc contour abnormality were derived from two types of reference points, real and theoretical corners, by using a manual segmentation procedure standardized during pilot work. This section of the thesis includes a summary of the main findings as well as a discussion of the studies' limitations and strengths, further research plans, associated challenges for the measurement properties of the measures proposed, and potential applications for new quantitative measures. The quantitative measures of disc contour abnormality proposed in this thesis, at this stage of their development, are recommended for research on the determinants and consequences of disc contour abnormalities and for longitudinal studies.

SUMMARY OF THE MAIN FINDINGS

Reliability

- Quantitative measures of lumbar disc contour abnormality have intra- and inter-evaluator reliability adequate for inferences at the group level[1] (73/80 generalizability coefficients estimates > 0.70), and superior to that of qualitative ratings by experienced medical specialists (Gcoefs 20/20 < 0.65).
- Inter-scanner reliability was adequate for group inferences (44/50 generalizability coefficients > 0.70)
- The effect of MR image resolution on the reliability of the quantitative measures of disc contour abnormality was smaller (2.0% of error variance) than expected relative to that due to intra- (1.2%-3.0%), inter-evaluator (5.0%), and inter-scanner differences (8.3%). Reliability results suggest that the same MR equipment, standardized imaging parameters, subject's positioning, and evaluator should be used when assessing quantitative disc contour abnormality, particularly when planning

follow-up or multi-center studies where a subject will possibly be repeatedly imaged on different scanners.

Effect of Position

- Posterior contour abnormalities were found to be significantly smaller on images obtained in flexion and in extension than in the neutral position (9.5%-30.1%). Anterior disc contour abnormalities were significantly smaller in extension than in flexion for all disc levels (17.2%-71.7%) and smaller in the neutral position than in flexion (26.3%-52.7%).
- Standardizing patients' positioning to assess the dimensions of disc contour abnormality is especially important when imaging subjects suspected of having important contour abnormalities because the differences between positions were found to be larger in subjects with larger contour abnormalities.

Convergent and Divergent Validity

The convergent and divergent validity of quantitative measures of disc contour abnormality was assessed by using qualitative ratings from different medical specialists and different rating scales representative of current clinical practice.

- The correlations between quantitative measures of contour abnormality and qualitative measures of bulging (anteriorly and posteriorly) were stronger (supporting the convergence hypothesis) than the weak correlations with qualitative endplate irregularity ratings (supporting the divergence hypothesis);
- The correlations among quantitative measures of disc contour abnormality on the same side of the disc (anteriorly or posteriorly) were stronger (supporting the convergence hypothesis) than among the quantitative measures of contour abnormality on the opposite side of the disc which were weak (supporting the divergence hypothesis).
- The pattern of the correlations between the quantitative contour abnormality measures and the qualitative ratings obtained on para-sagittal images also provided consistent evidence of convergent and divergent validity.

- Contrary to the study's hypothesis, the correlations between the quantitative measures of disc contour abnormality with the qualitative ratings of bulging from parasagittal images were lower rather than higher or equal to the correlations observed by using only midsagittal images.
- The lower convergent correlations observed with qualitative ratings, even if they were stronger than the divergent coefficients, can be explained by the low variance of the qualitative ratings in this sample and the lower reliability of the qualitative measures, which created attenuation of the correlation coefficients.

Differences in Contour Abnormalities Measurements between Midsagittal and Para-sagittal Images

- The midsagittal quantitative measurements of anterior and posterior disc contour abnormality were significantly smaller than measurements on at least three of four para-sagittal slices.
- In subjects with symmetrical moderate and large disc bulges a similar pattern of differences was observed. However, in this group more statistically significant interactions were found between the slice position and disc levels for anterior measures than in subjects without bulging. This finding demonstrates the capacity of the quantitative measures of contour abnormality to detect differences in the dimensions of contour abnormality in groups where differences are expected to be found.

STUDY LIMITATIONS

Distinguishing Bulging, Herniation, Protrusion and Extrusion Contained or not

The quantitative measures of disc contour abnormality assessed in the current study are adequate to quantify the amount of disc material located beyond the real or theoretical normal margins of the intervertebral disc space. However, the quantitative measures of disc contour abnormality assessed in this thesis do not allow differentiation between a bulge and various degrees of herniation according to the definitions in the nomenclature and classification of lumbar disc pathology recommended by the Combined Task Forces of the North American Spine Society, American Society of Spine Radiology, and American Society of Neuroradiology.[2] Those definitions are reported in Appendix B to support this discussion.

Distinguishing between bulging and herniation is not possible with the current quantitative measures of contour abnormality because, when measurements are obtained from midsagittal images, one cannot determine on what proportion of the disc's circumference the contour abnormality is present. Other quantitative measures derived from all sagittal and axial MR images could be designed to allow bulging to be distinguished from herniation, but such an approach would be very time-consuming unless automatic segmentation procedures became available. The distinction may be important because bulging may represent a pathologic finding, a physiologic variant, or normalcy.[2,3] However, previous studies using qualitative ratings have found that the prevalence of bulging and herniation is similar in subjects with and without low back problems.[4-7] Alternatively, the distinction between these basic definitions may not be as important as quantifying the amount of disc material located beyond the disc margins by assessing the size of the disc contour abnormality.

The current quantitative measures also do not assess whether in at least one plane, any one distance between the edges of the disc material beyond the disc space is greater than the distance between the edges of the base in the same plane. In other words, the measures assessed in this thesis do not distinguish between protrusion and extrusion. Quantitative measures including a ratio of the distance between the edges of the disc material located beyond the disc space to the distance between the edges of the base in the sagittal plane and in the axial plane could allow protruded material to be distinguished from extruded material. However, in light of new findings on the pathophysiology of discogenic pain, distinguishing between contained (Appendix B) and non-contained disc contour abnormalities may be more important because exposure to nucleus material leads to the inflammation of the surrounding tissues, especially in the presence of mechanical pressure.[8-21] However, the technical limitations of the currently available non-invasive imaging modalities (CT and MRI) usually preclude distinguishing between a contained and an uncontained disc contour abnormality.[2] As non-invasive imaging techniques allowing for the assessment of the integrity of the annulus, the longitudinal ligament, or the peridural membrane become available, new quantitative measures distinguishing contained from uncontained contour abnormality could be developed for use in etiologic research

Manual Segmentation is a Source of Error and Time-consuming

In order to obtain the quantitative measures of disc contour abnormality described in this thesis, the evaluators had to manually segment each MR image from which measurements were to be obtained. For an experienced evaluator, this process is time-consuming and takes approximately 15-20 minutes per sagittal slice. The training of an evaluator is also time-consuming. To become familiar with the segmentation procedure described in this thesis, a new and an experienced evaluator segmented together the images from four to five subjects representing a range of degenerative findings. The new evaluator observed the segmentation of the first two subjects by the experienced evaluator and then, completed the segmentation of the others while receiving feedback. The new evaluator practiced, segmenting the images from approximately 30 subjects that had previously been segmented by the experienced evaluator. Controversial cases were then discussed with the experienced evaluator and a new set of subjects were selected for additional training, if necessary, following that discussion. The new evaluator contributing to the intra- and inter-evaluator reliability estimates reported in this thesis had segmented the images of approximately 160 subjects prior to segmenting the images of the subjects that contributed to the measures. While possibly acceptable in research, the segmeation and training time are too long to be useful in a clinical environment. Such long segmentation times also reduce the feasibility of segmenting all sagittal images from given subjects in both the axial and sagittal planes, as would be required to distinguish bulging from herniation and protrusion from extrusion.

To my knowledge, automatic segmentation procedures, which could help accelerate segmentation, do not work well for segmenting disc tissue from the epidural tissue posteriorly (i.e., the ligament or the venous plexus) or from the blood vessels anteriorly (i.e., the vena cava, the aorta, or the iliac arteries).[22] Currently, such procedures are adequate only for bony structures and the spinal canal space.[23-27] Alternatively, accelerating the segmentation process by as much as 15% may be possible by using a stylus and a high-resolution touch-sensitive screen rather than a mouse and a conventional computer screen, as has been shown recently for the segmentation of knee cartilage.[28] Such an input strategy should be assessed by using SpEx©, the custom-designed spine analysis software used in this thesis, to determine if time savings can be realized and reliability maintained or improved.

The evaluators initially segmented the outlines of the disc and vertebrae on the proton-density images and then copied them onto the corresponding T1- and T2-weighted images to take advantage of the different contrasts. After pilot work, this approach was selected because the evaluators felt that it was needed in order to improve the anatomical relevance of the outlines. However, this approach may have introduced error into the measurement of contour abnormality because the patients can move between the acquisitions of the different sequences. This approach also increased the time required to acquire the images and complete the segmentation. With improvements in MR technology, obtaining an adequate segmentation in a shorter amount of time may become possible by using only a single sequence with an excellent contrast-to-noise ratio between the structures of interest. Nevertheless, the intra- (2%) and inter-evaluator (5%) errors for the quantitative measures of disc contour due to the standardized manual segmentation approach used in this thesis had only a relatively small negative effect on the reliability.

STUDY STRENGTHS

The development of the quantitative measures of disc contour abnormality proposed in the current thesis followed a rigorous approach. The current research considered four different types of measures unlike most previous studies on quantitative measures of disc contour abnormality which considered just a single type. The measures proposed were derived from a manual segmentation of the contours of the discs and vertebrae in order to define reference points from which to measure contour abnormality. This approach is believed to be more objective than directly measuring disc contour abnormality on the images, with an evaluator placing the cursor where he perceives the reference points are. The manual segmentation procedure was standardized following the inspection of outliers in repeated disc contour abnormality measurements obtained in a pilot study of 79 subjects. The segmented outline was created by using the information from three MR sequences routinely acquired: Proton density, T1-, and T2 weighted images. The procedure takes advantage of the signal intensity differences between the structures of interest to refine the outline from which the measures are derived even though the quantitative measures of disc abnormality are of distance and areas and not of signal intensity.

The reliability of each of the four different quantitative measures was compared to that of the others and to that of qualitative measures obtained in the same sample of images with a sufficient sample size by using a strong statistical approach: generalizability theory. The main advantage of generalizability theory is its flexibility in examining the effect of more than one factor at a time on measurement reliability. Furthermore, G-theory allows for the quantification of the proportion of the measurement error attributable to each specific facet and also allows for the estimation of the reliability that could be attained by averaging repeated measurements to control the measurement error due to a specific facet. In contrast, the external validity of previous reliability studies of quantitative measures of disc contour abnormality is questionable due to methodological problems controlled for in the current study (sampling,[29,30] statistics used,[30] and dichotomizing of a continuous measure[31]). The evaluators in the current research were blind to each others' work and to the subjects' clinical information and identities.

In addition, to my knowledge, no other quantitative measure of disc contour abnormality has been assessed for as many potential relevant sources of errors as the quantitative measures proposed in the current study. The factors influencing the reliability of other quantitative measures of disc contour abnormality from MR images studied to date include intra-[29-33] and inter-evaluator difference.[29,30,33] Only two other studies, with inadequate sample size (n = 1), had estimated scan-rescan reliability for measures of disc contour abnormality,[30,33] and none had examined the impact of image-resolution differences. Studies of the influence of inter-scanner differences on measurement properties for the lumbar spine are rare, and none included direct measures of disc contour abnormality.[34-36] In contrast, the influence of intra-, inter-evaluator, inter-resolution, inter-scanner, and inter-position differences was quantified for the measures in the present studies.

Studies of the validity of the magnitude assessment of quantitative disc contour abnormality are rare. To my knowledge, previous validity studies of quantitative measures have considered only the correlation between different types of other quantitative measures obtained by the same evaluators.[29,37,38] In addition, in this thesis, convergent and divergent validity was also studied with the current standard of clinical assessment of MR images, qualitative ratings by independent raters, and provided further validity evidence. Many efforts were made to improve the external generalizability of the validity findings. Analyses were repeated by using ratings from three different evaluators, using three sets of rating definitions representative of definitions commonly used in practice and using data extracted from different images. Furthermore, the samples sizes were larger than those in previous validity studies. Nevertheless, while validity evidence was produced in the current study, the validation process should not be considered complete at this stage. Stronger validity evidence could be provided by using the study designs discussed in the validity chapter.

FURTHER DEVELOPMENT WORK SUGGESTED FOR QUANTITATIVE MEASURES OF CONTOUR ABNORMALITY

Exploring Additional Sources of Unexplained Measurement Error

Most of the measurement error variance remained unaccounted for even after quantifying the proportion of measurement error in the quantitative measures of disc contour abnormality attributed to intra-, inter-evaluator, inter-scanner, and inter-resolution differences. Additional studies could be conducted to study the impact of the following factors on the reliability of quantitative measures of disc contour abnormality, as they have been shown to influence the reliability of other quantitative measures: acquisition factors (repositioning in the same scanner,[39-54] MR sequences,[53,55] image signal-to-noise ratio,[39] MR field strength,[35,36,41,56] slice thickness[34,41,48,56]), subjects' factors (pathological vs. healthy[42,57]), and image analysis factors (mouse vs. digitizing tablet,[58] digital displays,[58] digitizer,[39,41,42,44] and segmentation technique[39,42-44,49,51,57]).

Such studies may help suggest standardization procedures to maximize the reliability of a single quantitative measure of contour abnormality and reach reliability levels adequate for inferences at the individual level.[1,59]

Image Segmentation Procedure

The standardized manual segmentation procedure used to obtain the quantitative measures of disc contour abnormality in this thesis allowed for measurements with reliability adequate for group-level inferences.[1] However, the use of new technology for evaluator input in manually segmenting the images may also lead to better reliability. Alternatively, automatic segmentation processes have been developed in research on other anatomical structures and could eventually be applied to facilitate the segmentation of soft tissues on lumbar images. Automatic segmentation may also become possible if image quality and the contrast-to-noise ratio improve with new imaging technology or sequences. The impact of such new developments on the reliability of quantitative measures of contour abnormality should be studied.

For measures of disc contour abnormality that use theoretical corners, the position of the theoretical vertebra corners was defined by the intersection of a line connecting the diameter of the vertebra above the disc at 20% of the vertebra height to the corresponding diameter below the disc with lines representing the disc-vertebra boundaries above and below the disc. A study of the influence of using a different percentage of the vertebral height to define the position of the theoretical vertebral corners on the reliability of quantitative measures of disc contour abnormality is underway. The current selection of a proportion of 20% of the vertebra height was made on the basis that this proportion allowed for the use of a reference point on a part of the vertebra's contour that could reliably be traced. That position of the vertebra contour was considered less likely to be influenced by the presence of osteophytes, which typically grow closer to the endplates. Further, that position was thought to be more easily outlined during manual segmentation than a point at mid-height, which Harada et al.[60] used, because on MR images, the vertebral artery enters the vertebral body at mid-height. However, the presence of the vertebral artery, epidural veins, and fat along the posterior wall of the vertebra makes reliably segmenting the contour of the vertebral wall from these tissues challenging.

Segmentation of this area is adjusted by using all three (PD, T1-, T2-) sequences because the artery is best seen on T2-, the epidural fat and bone are best seen on T1-, and the posterior longitudinal ligament is best seen on PD.

Providing Stronger Validity Evidence

A criterion validity study could be conducted to obtain disc contour abnormality measurements from the sagittal anatomical and MRI slices by using frozen fresh human cadavers or animal spines, as has been done for brain images.[61] In such a study, staining anatomical slices may facilitate the segmentation process and reduce the measurement error of disc contour due to the manual segmentation of anatomical slices. Alternatively, criterion validity could be assessed *in vivo* in subjects operated on for disc contour abnormality. Measures of quantitative disc contour abnormality from MR images obtained before surgery could be compared with the quantification of contour abnormality during surgery or the quantification of the amount of tissue removed during surgery. However, different surgeons assess surgical findings differently and the amount of tissue removed does not necessarily correspond to that seen on MR images beyond the interspace. The criterion validity of the quantitative measurement procedures for disc contour abnormality could also be examined by using imaging phantoms created to have known dimensions of contour abnormality in the planes examined with MRI, as has been done in brain and cartilage research.[27,39,49,54] Concurrent validity could be examined by deriving quantitative contour abnormality measures from images of the same subject by using different imaging techniques, as has been done for femoral bone geometry by using MRI and CT scans.[62] Finally, as in the development of segmentation procedures for brain images,[57] an alternative validity study could compare the quantitative contour abnormality measures derived with the procedure described using SpEx© to measures derived by using other segmentation procedures and measurement software.

Which Slice(s) should Be Used to Measure Contour Abnormality Quantitatively?

In the chapter on validity, midsagittal images were shown to lead to smaller contour abnormality measurements than para-sagittal images (right and left). Future research is needed to determine whether quantitative contour abnormalities should be measured on all sagittal images or whether a subset of images is sufficient to provide a representative assessment. Another alternative to using a limited number of sagittal slices to obtain a disc contour abnormality measurement is to compute a volume measurement by combining the quantitative measures of disc contour abnormality from all sagittal slices.[38,63] This method's advantage is that no information about the size of the disc contour abnormality would be lost. Future research comparing the validity and reliability of different approaches will assist in selecting which to use for studying the relationships between quantitative disc contour abnormality and pain or impairment measures.

Study Alternative Measures of Disc Contour Abnormality

Quantitative measures including the ratio of the distance between the edges of the disc material beyond the disc space to the distance between the edges of the base in the sagittal plane and in the axial plane could be designed and evaluated to allow protruded material to be distinguished from extruded material. Such a ratio could be obtained in both the axial and sagittal planes to fully characterize disc contour abnormality in a manner consistent with the definition from the consensus of the leading radiological societies.[2] Alternatively, if the segmentation process could be accelerated, measures of the volume of disc material located beyond the disc space could also be obtained. While the latter measure would provide a full assessment of volume, it would not allow evaluators to distinguish between bulging and herniation, and protrusion and extrusion. Combinations of the two measures described above may provide clinically relevant information on disc contour abnormality.

Other quantitative measures could also be designed to more fully represent whether the presence of disc material beyond the disc space boundaries is acting as a space-occupying lesion compromising the reserve space surrounding the nerve tissue. Evidence suggests that compromise of the spinal cord or nerve roots is more strongly related to pain than the absolute magnitude of disc contour abnormality.[5,7,64-66] However, to my knowledge, no quantitative measures of nerve root compromise have been assessed for reliability in the sagittal plane. The measures proposed in this thesis could be used as reliable measures for the numerator in quantitative ratios indicating nerve compromise due to disc contour abnormality. Denominators to consider for such ratios could include the spinal canal diameter, the distance between the disc boundary and the nerve root or the spinal cord, or the remaining reserve space for the neural tissue. Measures of the denominator will be designed and assessed for reliability by a collaborator from our research team. The ratios of disc contour abnormality to the remaining canal or foramen space (including nerve tissue) in the axial plane have been assessed for reliability by other research groups.[32,33,67,68] However, the distances from the canal, foramen boundaries, or the disc to the nerve tissue itself have not been assessed specifically in either the sagittal or axial plane.

POTENTIAL APPLICATIONS FOR QUANTITATIVE MEASURES OF DISC CONTOUR ABNORMALITY

Longitudinal Studies

Mounting evidence suggests that reductions in the size of the focal morphological disc findings occur as a result of natural history and conservative or surgical interventions.[69-76] Reliable quantitative measures of disc contour abnormality will help reduce the sample size needed to quantify such changes.[74,76] Quantitative measures of disc contour abnormality are already being used for such studies.[32,33,67,76-81] Alternatively, the quantitative measures of contour abnormality proposed in this thesis could also be useful in studying whether new pharmacological treatments have a negative impact on the natural resorption of disc material within the disc contour abnormality.[82,83] Similarly, quantitative measures would be useful to determine whether recent advances in biological or genetic interventions are successful at preventing the progression of degenerative disc disease.[84-86] Quantitative measures have more increments and are likely to be more responsive to change than qualitative measures and, thus, to be more useful in longitudinal studies where small changes are of interest. Also, when no change is expected, a more sensitive measure is less susceptible to a Type 2 error, or the conclusion that no difference exists when in fact one does.[87(p262)]

Longitudinal studies could also evaluate the predictive value of baseline quantitative measures of contour abnormality for outcomes of a given treatment approach. Prognostic studies of the appearance of focal disc contour abnormalities suggest that the largest contour abnormalities, with enhanced signal intensity after gadolinium injection or non-contained/sequestred focal disc contour abnormalities, are the most likely to diminish in size over time without surgical intervention.[71,74,75,78,88] Such findings could be confirmed by using the quantitative measures of contour abnormality studied in this thesis. Follow-up studies of disc contour abnormality using the quantitative measures studied in this thesis are currently underway.

Studies on the Determinants and Epidemiology of Disc Contour Abnormality

Quantitative measures were shown to have superior reliability and to measure with more increments, particularly for smaller contour abnormalities, than qualitative ratings. The higher number of increments using quantitative contour abnormality measures resulted in wider measurement variance for a given sample, as reported previously in the chapter five. Qualitative measures of bulging appear to be characterized by a floor effect in population samples while quantitative measures are able to distinguish between smaller degrees of contour abnormality. The study of the determinants of disc contour abnormality is usually conducted by using population samples in which very large contour abnormalities are less common. Quantitative measures will likely be useful in correlational studies to identify the determinants of disc contour abnormality in such samples. However, prior to conducting such a study, determining which data to use in the correlations would be necessary. While a summary midsagital measure of disc contour abnormality may be adequate to study determinants suspected to have a systemic (general) effect, such a measure may not be useful for a determinant suspected to affect a specific location of a given disc (e.g., the posterolateral aspect of L4L5).

In a recent review of studies on the epidemiology and genetic influence of lumbar disc degeneration, our research group observed that such studies used mostly qualitative and unstandardized measures of disc degeneration.[6] We concluded that: "There is no standard definition of disc degeneration; therefore, determinations of its presence and severity vary widely and are determined largely by the technology used to interrogate the disc. The challenge of developing more uniform case definitions of disc degeneration and accurate, reliable measures needs to be addressed to further advance studies in this area."[6] This recommendation is supported by the observation, using another quantitative measure, of a significant association between a measure of disc signal intensity and two polymorphisms of the vitamin D receptor gene which was revealed when the quantitative measure was used but was undetected by using a less reliable qualitative grading scale.[89-91] The development of quantitative measures of disc contour abnormality is another step in the right direction. The influence of genetic and environmental factors, such as occupational or leisure activity loading, on the progression of disc contour abnormality over time is currently being studied by using quantitative data from the large epidemiological study from which many of the study samples presented in this thesis were drawn.

Quantitative measures will also help clarify whether disc contour abnormality is strongly associated with other measures of disc degeneration. Previous studies using qualitative ratings or unstandardized measures have shown strong correlations amongst measures of different degeneration signs.[38,60,92] A common underlying pathological process may be leading to the progression of different degeneration measurements. Quantitative assessment of multiple signs of degeneration may help verify the hypothesis above.

Studies of the Effect of Postures and Loading on Disc Contour Abnormality

The objective of some of the early studies using quantitative measurements was to measure for diurnal variation in the degree of disc contour abnormality[93] and to determine whether such variation could be related to spine loading.[30] Further to this, patients often report that specific postures influence their symptoms. Quantitative measures could be used to determine if those postures have an impact on the dimension of disc contour abnormality and/or on the relationship between contour abnormalities and the nerve roots. Moreover, some exercise approaches are based on the theory that repeated movements or sustained end-range postures in specific directions have an impact on the intervertebral disc including, possibly, disc contour abnormalities.[94-98] The quantitative measures studied in this thesis could be used to evaluate this theory.

A study of the correlations between changes in the degree of contour abnormalities, disc height, and the position of the nucleus based on signal intensity analysis will be undertaken. Such a study will provide insights into the mechanism underlying the changes measured in the chapter on the effect of flexion and extension on quantitative measures of disc contour abnormality. The study of the correlations between changes in the dimensions of disc contour abnormalities, spinal canal stenosis, foramen stenosis, nerve tissue thickness, and ligament thickness would also be of interest. Such a study could help identify the determining factors for the increasing foramen and canal stenosis occurring when the spine moves toward extension. An alternative explanation is needed because the hypothesis that an increase in stenosis in extension is due to increased contour abnormality[30,64,99,100] was not supported by the results presented in Chapter 4.

Studies of the Signs and Symptom, if any, Associated with Disc Contour Abnormality

The quantitative measures presented in this thesis are appropriate for studying whether disc contour abnormality is a cause of low back pain via a space-occupying mechanism. However, in the last two decades, progress in the understanding of the pathophysiology of discogenic low back pain has shown that mechanical compression is not the only mechanism by which the disc can cause LBP.[9-21] The current quantitative measures, like qualitative ratings of bulging or herniation severity, may not correlate with discogenic pain arising from inflammation, immunological, or chemical irritation of nerve endings due to leakage of disc content unless the intensity of such reactions is proportional to the exposed disc material.

However, in cases where pain arises due to the irritation of the nerve endings in the outer third of the annulus fibrosus by nuclear material, an increase of the signal intensity in the areas of disc contour abnormality defined in this study could possibly correlate with the symptoms of LBP. The signal intensity of the nucleus material is higher than that of the fibrous tissue in the annulus fibrosus, and the presence of nuclear material in the annulus could be detected by an elevation of signal intensity in the annulus fibrosus area. Alternatively, the contrast enhancement of areas where nerve ingrowth occurs in annular tears infiltrated by disc material has been described in cases of chronic low back pain, following the injection of gadolinium.[66,101-103] Quantifying the increased signal intensity within the areas of contour abnormality defined in this thesis may also be a useful method of detecting areas of nerve and vessel ingrowth in the disc annulus or predicting which contour abnormalities are the most likely to resorb over time.[74,101-104] Non-contained disc material enhancing when gadolinium is injected has been shown to be more likely to resorb. The measurement properties of such an application of the areas defined in this thesis should be assessed.

Further study of the correlations between the signs, symptoms, and quantitative measures of contour abnormality discussed in the current study would help clarify whether low/unknown reliability and range restriction are factors attenuating the correlations found to date by using qualitative bulging ratings.[37,105,106]

CONCLUSION

As part of this dissertation, quantitative measures of disc contour abnormality were developed by consensus. As expected, reliability studies demonstrated that quantitative measures of disc contour abnormality are more reliable than the qualitative ratings by clinicians and adequate for inferences at the group level. However, the reliabilities of a single quantitative measurement are not high enough for inferences at the individual level. The sources of errors in the quantitative measures were explored and showed that image resolution, intra-evaluator, inter-evaluator, MR scanner, and imaging positions differences could influence these measures. Evidence of the construct validity of the new quantitative measures of disc contour abnormality was also provided.

The development of outcome measures is an ongoing process. As a result, further validity assessment of these new quantitative measures of contour abnormality is needed. The impact of other sources of error on the reliability of the quantitative measures should also be studied when relevant for specific research projects. Finally, the predictive value of quantitative measures of contour abnormality in guiding clinical decision needs to be assessed before these measures are used with patients.

At this stage of their development, the quantitative measures are recommended for research on the determinants and effects of disc contour abnormalities. The improved reliability will likely assist in detecting associations that could not be highlighted by using qualitative ratings with inferior reliability and limited measurement increments. The quantitative measures use a continuous scale providing more increments. It promises to be useful for longitudinal studies in the assessment of the effect of natural history and conservative, surgical, biological, or genetic interventions on disc contour abnormalities.

REFERENCES - CHAPTER 6

1. Scientific Advisory Committee of the Medical Outcome Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res.* 2002;11(3):193-205.
2. Fardon DF, Milette PC. Nomenclature and classification of lumbar disc pathology. Recommendations of the Combined task Forces of the North American Spine Society, American Society of Spine Radiology, and American Society of Neuroradiology. *Spine.* 2001;26(5):E93-E113.
3. Porter RW. Does hard work prevent disc protrusion? *Clin Biomech.* 1987;2(4):196-198.
4. Jensen MC, Brant-Zawadzki MN, Obuchowski N et al. Magnetic resonance imaging of the lumbar spine in people without back pain. *N Engl J Med.* 1994;331(2):69-73.
5. Boos N, Rieder R, Schade V et al. 1995 Volvo Award in clinical sciences. The diagnostic accuracy of magnetic resonance imaging, work perception, and psychosocial factors in identifying symptomatic disc herniations. *Spine.* 1995;20(24):2613-2625.
6. Battie MC, Videman T, Parent E. Lumbar Disc Degeneration: Epidemiology and Genetic Influences. *Spine.* 2004;29(23):2679-2690.
7. Jarvik JJ, Hollingworth W, Heagerty P et al. The longitudinal assessment of imaging and disability of the back (laidback) study: baseline data. *Spine.* 2001;26(10):1158-1166.
8. Videman T, Battie MC, Gibbons LE et al. Associations between back pain history and lumbar MRI findings. *Spine.* 2003;28(6):582-588.
9. McCarron RF, Wimpee MW, Hudkins PG et al. The inflammatory effect of nucleus pulposus. A possible element in the pathogenesis of low-back pain. *Spine.* 1987;12(8):760-764.
10. Olmarker K, Blomquist J, Stromberg J et al. Inflammatogenic properties of nucleus pulposus. *Spine.* 1995;20(6):665-669.
11. Garfin SR, Rydevik BL, Brown RA. Compressive neuropathy of spinal nerve roots. A mechanical or biological problem? *Spine.* 1991;16(2):162-166.
12. Iwabuchi M, Rydevik B, Kikuchi S et al. Effects of anulus fibrosus and experimentally degenerated nucleus pulposus on nerve root conduction velocity: relevance of previous experimental investigations using normal nucleus pulposus. *Spine.* 2001;26(15):1651-1655.

13. Harrington JF, Messier AA, Bereiter D et al. Herniated lumbar disc material as a source of free glutamate available to affect pain signals through the dorsal root ganglion. *Spine.* 2000;25(8):929-936.
14. Nygaard OP, Mellgren SI, Osterud B. The inflammatory properties of contained and noncontained lumbar disc herniation. *Spine.* 1997;22(21):2484-2488.
15. Olmarker K, Nordborg C, Larsson K et al. Ultrastructural changes in spinal nerve roots induced by autologous nucleus pulposus. *Spine.* 1996;21(4):411-414.
16. Marshall LL, Trethewie ER, Curtain CC. Chemical radiculitis. A clinical, physiological and immunological study. *Clin Orthop.* 1977;(129):61-67.
17. Gertzbein SD, Tile M, Gross A et al. Autoimmunity in degenerative disc disease of the lumbar spine. *Orthop Clin North Am.* 1975;6(1):67-73.
18. Videman T, Nurminen M. The occurrence of anular tears and their relation to lifetime back pain history: a cadaveric study using barium sulfate discography. *Spine.* 2004;29(23):2668-2676.
19. Olmarker K. Mechanical and biochemical injury of spinal nerve roots: an experimental perspective. In: Weinstein JN, Gordon SL, editors. *Low back pain A scientific and clinical overview.* Rosemont, Illinois: American Academy of Orthopaedic Surgeons, 1996: 215-233.
20. Marshall LL, Trethewie ER. Chemical irritation of nerve-root in disc prolapse. *Lancet.* 1973;2(7824):320.
21. Saal JS. The role of inflammation in lumbar pain. *Spine.* 1995;20(16):1821-1827.
22. Ramirez L, Parent E, Durdle NG et al. Interactive Segmentation for Quantitative Analysis of Disc and Vertebral Morphometric Characteristics". *International Research Society for Spinal Deformities.* 2004;45.
23. Benameur S, Mignotte M, Parent S et al. 3D/2D registration and segmentation of scoliotic vertebrae using statistical models. *Comput Med Imaging Graph.* 2003;27(5):321-337.
24. Hoad CL, Martel AL. Segmentation of MR images for computer-assisted surgery of the lumbar spine. *Phys Med Biol.* 2002;47(19):3503-3517.
25. Huesman RH, Klein GJ, Kimdon JA et al. Deformable registration of multimodal data including rigid structures. *IEEE Transactions on Nuclear Science.* 2003;50(3):389-392.
26. Penning L, Irwan R, Oudkerk M. Measurement of angular and linear segmental lumbar spine flexion-extension motion by means of image registration. *Eur Spine J.* 2005;14(2):163-170.
27. Rogers BP, Haughton VM, Arfanakis K et al. Application of image registration to measurement of intervertebral rotation in the lumbar spine. *Magn Reson Med.* 2002;48(6):1072-1075.
28. McWalter EJ, Wirth W, Siebert M et al. Use of novel interactive input devices for segmentation of articular cartilage from magnetic resonance images. *Osteoarthritis Cartilage.* 2005;13(1):48-53.
29. Cooley JR, Danielson CD, Schultz GD et al. Posterior disk displacement: morphologic assessment and measurement reliability-lumbar spine. *J Manipulative Physiol Ther.* 2001;24(5):317-326.
30. Fredericson M, Lee SU, Welsh J et al. Changes in posterior disc bulging and intervertebral foraminal size associated with flexion-extension movement: a

comparison between L4-5 and L5-S1 levels in normal subjects. *Spine J.* 2001;1(1):10-17.

31. Luoma K, Vehmas T, Raininko R et al. Lumbosacral transitional vertebra: relation to disc degeneration and low back pain. *Spine.* 2004;29(2):200-205.
32. Buttermann GR. Lumbar disc herniation regression after successful epidural steroid injection. *J Spinal Disord Tech.* 2002;15(6):469-476.
33. Wood KB, Blair JM, Aepple DM et al. The natural history of asymptomatic thoracic disc herniations. *Spine.* 1997;22(5):525-529.
34. Greenstein J, Cramer GD, Howe J et al. Comparison of 1.5 Tesla and 0.35 Tesla field strength magnetic resonance imaging scans in the morphometric evaluation of the lumbar intervertebral foramina. *J Manipulative Physiol Ther.* 1995;18(4):195-202.
35. Merl T, Scholz M, Gerhardt P et al. Results of a prospective multicenter study for evaluation of the diagnostic quality of an open whole-body low-field MRI unit. A comparison with high-field MRI measured by the applicable gold standard. *Eur J Radiol.* 1999;30(1):43-53.
36. Cihangiroglu M, Yildirim H, Bozgeyik Z et al. Observer variability based on the strength of MR scanners in the assessment of lumbar degenerative disc disease. *Eur J Radiol.* 2004;51(3):202-208.
37. Thelander U, Fagerlund M, Friberg S et al. Describing the size of lumbar disc herniations using computed tomography. A comparison of different size index calculations and their relation to sciatica. *Spine.* 1994;19(17):1979-1984.
38. Holodny AI, Kisza PS, Contractor S et al. Does a herniated nucleus pulposus contribute significantly to a decrease in height of the intervertebral disc? Quantitative volumetric MRI. *Neuroradiology.* 2000;42(6):451-454.
39. Gougoutas AJ, Wheaton AJ, Borthakur A et al. Cartilage volume quantification via Live Wire segmentation. *Acad Radiol.* 2004;11(12):1389-1395.
40. Schaefer SM, Abercrombie HC, Lindgren KA et al. Six-month test-retest reliability of MRI-defined PET measures of regional cerebral glucose metabolic rate in selected subcortical structures. *Hum Brain Mapp.* 2000;10(1):1-9.
41. Bird P, Ejbjerg B, McQueen F et al. OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging Studies. Exercise 5: an international multicenter reliability study using computerized MRI erosion volume measurements. *J Rheumatol.* 2003;30(6):1380-1384.
42. Glaser C, Burgkart R, Kutschera A et al. Femoro-tibial cartilage metrics from coronal MR image data: Technique, test-retest reproducibility, and findings in osteoarthritis. *Magn Reson Med.* 2003;50(6):1229-1236.
43. Vemuri P, Kholmovski EG, Goodrich KC et al. Statistics-based approach for aneurysm volume measurements. *J Magn Reson Imaging.* 2004;20(2):340-346.
44. Wang Q, Robson MD, Francis JM et al. Accuracy of quantitative MR vessel wall imaging applying a semi-automated gradient detection algorithm--a validation study. *J Cardiovasc Magn Reson.* 2004;6(4):895-907.
45. Westwood MA, Anderson LJ, Firmin DN et al. Interscanner reproducibility of cardiovascular magnetic resonance T2* measurements of tissue iron in thalassemia. *J Magn Reson Imaging.* 2003;18(5):616-620.

46. Firbank MJ, Coulthard A. Evaluation of a technique for estimation of extraocular muscle volume using 2D MRI. *Br J Radiol.* 2000;73(876):1282-1289.
47. Mahrholdt H, Wagner A, Holly TA et al. Reproducibility of chronic infarct size measurement by contrast-enhanced magnetic resonance imaging. *Circulation.* 2002;106(18):2322-2327.
48. Filippi M, Marciano N, Capra R et al. The effect of imprecise repositioning on lesion volume measurements in patients with multiple sclerosis. *Neurology.* 1997;49(1):274-276.
49. Tan IL, van Schijndel RA, van Walderveen MA et al. Magnetic resonance image registration in multiple sclerosis: comparison with repositioning error and observer-based variability. *J Magn Reson Imaging.* 2002;15(5):505-510.
50. Giedd JN, Kozuch P, Kaysen D et al. Reliability of cerebral measures in repeated examinations with magnetic resonance imaging. *Psychiatry Res.* 1995;61(2):113-119.
51. Crum WR, Scahill RI, Fox NC. Automated hippocampal segmentation by regional fluid registration of serial MRI: validation and application in Alzheimer's disease. *Neuroimage.* 2001;13(5):847-855.
52. Rovaris M, Gawne-Cain ML, Sormani MP et al. The effect of repositioning on brain MRI lesion load assessment in multiple sclerosis: reliability of subjective quality criteria. *J Neurol.* 1998;245(5):273-275.
53. Steens SC, Admiraal-Behloul F, Schaap JA et al. Reproducibility of brain ADC histograms. *Eur Radiol.* 2004;14(3):425-430.
54. Kovacevic N, Lobaugh NJ, Bronskill MJ et al. A robust method for extraction and automatic segmentation of brain images. *Neuroimage.* 2002;17(3):1087-1100.
55. Filippi M, Rocca MA, Gasperini C et al. Interscanner variation in brain MR lesion load measurements in multiple sclerosis using conventional spin-echo, rapid relaxation-enhanced, and fast-FLAIR sequences. *AJNR Am J Neuroradiol.* 1999;20(1):133-137.
56. Filippi M, van Waesberghe JH, Horsfield MA et al. Interscanner variation in brain MRI lesion load measurements in MS: implications for clinical trials. *Neurology.* 1997;49(2):371-377.
57. Wei X, Warfield SK, Zou KH et al. Quantitative analysis of MRI signal abnormalities of brain white matter with high reproducibility and accuracy. *J Magn Reson Imaging.* 2002;15(2):203-209.
58. McWalter E, Wirth W, Siebert M et al. Time Savings Using Interactive Input Devices for Cartilage Segmentation. *5th Combined Meeting of the Orthopaedic Research Societies of Canada, U.S.A., Japan and Europe.* 2005;68.
59. Streiner DL, Norman GR. *Health measurement scales: A practical guide to their development and use.* Second ed. New York, New York: Oxford Medical Publications, 1995;1-231.
60. Harada A, Okuizumi H, Miyagi N et al. Correlation between bone mineral density and intervertebral disc degeneration. *Spine.* 1998;23(8):857-861.
61. Bartlett TQ, Vannier MW, McKeel DW, Jr. et al. Interactive segmentation of cerebral gray matter, white matter, and CSF: photographic and MR images. *Comput Med Imaging Graph.* 1994;18(6):449-460.

62. Woodhead HJ, Kemp AF, Blimkie CJR et al. Measurement of midfemoral shaft geometry: repeatability and accuracy using magnetic resonance imaging and dual-energy X-ray absorptiometry. *J Bone Miner Res.* 2001;16(12):2251-2259.
63. Roberts N, Hogg D, Whitehouse GH et al. Quantitative analysis of diurnal variation in volume and water content of lumbar intervertebral discs. *Clin Anat.* 1998;11(1):1-8.
64. Weishaupt D, Schmid MR, Zanetti M et al. Positional MR imaging of the lumbar spine: does it demonstrate nerve root compromise not visible at conventional MR imaging? *Radiology.* 2000;215(1):247-253.
65. Pfirrmann CW, Dora C, Schmid MR et al. MR image-based grading of lumbar nerve root compromise due to disk herniation: reliability study with surgical correlation. *Radiology.* 2004;230(2):583-588.
66. Stadnik TW, Lee RR, Coen HL et al. Annular tears and disk herniation: prevalence and contrast enhancement on MR images in the absence of low back pain or sciatica. *Radiology.* 1998;206(1):49-55.
67. Tonami H, Yokota H, Nakagawa T et al. Percutaneous laser discectomy: MR findings within the first 24 hours after treatment and their relationship to clinical outcome. *Clin Radiol.* 1997;52(12):938-944.
68. Carragee EJ, Kim DH. A prospective analysis of magnetic resonance imaging findings in patients with sciatica and lumbar disc herniation. Correlation of outcomes with disc fragment and canal morphology. *Spine.* 1997;22(14):1650-1660.
69. Delauche-Cavallier MC, Budet C, Laredo JD et al. Lumbar disc herniation. Computed tomography scan changes after conservative treatment of nerve root compression. *Spine.* 1992;17(8):927-933.
70. Teplick JG, Haskin ME. Spontaneous regression of herniated nucleus pulposus. *AJR Am J Roentgenol.* 1985;145(2):371-375.
71. Komori H, Shinomiya K, Nakai O et al. The natural history of herniated nucleus pulposus with radiculopathy. *Spine.* 1996;21(2):225-229.
72. Saal JA, Saal JS, Herzog RJ. The natural history of lumbar intervertebral disc extrusions treated nonoperatively. *Spine.* 1990;15(7):683-686.
73. Ellenberg MR, Ross ML, Honet JC et al. Prospective evaluation of the course of disc herniations in patients with proven radiculopathy. *Arch Phys Med Rehabil.* 1993;74(1):3-8.
74. Gallucci M, Bozzao A, Orlandi B et al. Does postcontrast MR enhancement in lumbar disk herniation have prognostic value? *J Comput Assist Tomogr.* 1995;19(1):34-38.
75. Kawaji Y, Uchiyama S, Yagi E. Three-dimensional evaluation of lumbar disc hernia and prediction of absorption by enhanced MRI. *J Orthop Sci.* 2001;6(6):498-502.
76. Matsubara Y, Kato F, Mimatsu K et al. Serial changes on MRI in lumbar disc herniations treated conservatively. *Neuroradiology.* 1995;37(5):378-383.
77. Nakano M, Matsui H, Ishihara H et al. Serial changes of herniated intervertebral discs after posterior lumbar discectomy: the relation between magnetic resonance imaging of the postoperative intervertebral discs and clinical outcome. *J Spinal Disord.* 2001;14(4):293-300.

78. Henmi T, Sairyo K, Nakano S et al. Natural history of extruded lumbar intervertebral disc herniation. *J Med Invest.* 2002;49(1-2):40-43.
79. Bozzao A, Gallucci M, Masciocchi C et al. Lumbar disk herniation: MR imaging assessment of natural history in patients treated without surgery. *Radiology.* 1992;185(1):135-141.
80. Sahlstrand T, Lonntoft M. A prospective study of preoperative and postoperative sequential magnetic resonance imaging and early clinical outcome in automated percutaneous lumbar discectomy. *J Spinal Disord.* 1999;12(5):368-374.
81. Kato F, Mimatsu K, Kawakami N et al. Serial changes observed by magnetic resonance imaging in the intervertebral disc after chemonucleolysis. A consideration of the mechanism of chemonucleolysis. *Spine.* 1992;17(8):934-939.
82. Korhonen T, Karppinen J, Malmivaara A et al. Efficacy of infliximab for disc herniation-induced sciatica: one-year follow-up. *Spine.* 2004;29(19):2115-2119.
83. Autio R, Karppinen J, Niinimaki J et al. The Effect of infliximab, a monoclonal antibody against TNF-alpha, on disc herniation resorption, a randomized controlled trial. *Abstracts from the 32nd Annual Meeting of the International Society for the Study of the Lumbar Spine.* 2005;41.
84. Yoon ST. The potential of gene therapy for the treatment of disc degeneration. *Orthop Clin North Am.* 2004;35(1):95-100.
85. Gruber HE, Hanley EN, Jr. Recent advances in disc cell biology. *Spine.* 2003;28(2):186-193.
86. Hidaka C, Khan SN, Farmer JC et al. Gene therapy for spinal applications. *Orthop Clin North Am.* 2002;33(2):439-446.
87. Glass GV, Hopkins KD. *Statistical methods in education and psychology.* 3 ed. Needham Heights, MA, USA: Allyn and Bacon, 1995;1-674.
88. Ahn SH, Park HW, Byun WM et al. Comparison of clinical outcomes and natural morphologic changes between sequestered and large central extruded disc herniations. *Yonsei Med J.* 2002;43(3):283-290.
89. Battie MC, Videman T, Gibbons LE et al. 1995 Volvo Award in clinical sciences. Determinants of lumbar disc degeneration. A study relating lifetime exposures and magnetic resonance imaging findings in identical twins. *Spine.* 1995;20(24):2601-2612.
90. Videman T, Leppavuori J, Kaprio J et al. Intragenic polymorphisms of the vitamin D receptor gene associated with intervertebral disc degeneration. *Spine.* 1998;23(23):2477-2485.
91. Videman T, Nummi P, Battie MC et al. Digital assessment of MRI for lumbar disc desiccation. A comparison of digital versus subjective assessments and digital intensity profiles versus discogram and macroanatomic findings. *Spine.* 1994;19(2):192-198.
92. Videman T, Battie MC, Gill K et al. Magnetic resonance imaging findings and their relationships in the thoracic and lumbar spine. Insights into the etiopathogenesis of spinal degeneration. *Spine.* 1995;20(8):928-935.
93. Silcox DH, III, Horton WC, Silverstein AM. MRI of lumbar intervertebral discs. Diurnal variations in signal intensities. *Spine.* 1995;20(7):807-811.
94. McKenzie R, May S. *The Lumbar Spine Mechanical Diagnosis & Therapy.* 2nd ed. Waikanae, New Zeland: Spinal Publications New Zeland Ltd, 2003;1-374.

95. Wetzel FT, Donelson R. The role of repeated end-range/pain response assessment in the management of symptomatic lumbar discs. *Spine J.* 2003;3(2):146-154.
96. Long A, Donelson R, Fung T. Does it Matter Which Exercise?: A Randomized Control Trial of Exercise for Low Back Pain. *Spine.* 2004;29(23):2593-2602.
97. Donelson R, Aprill C, Medcalf R et al. A prospective study of centralization of lumbar and referred pain. A predictor of symptomatic discs and anular competence. *Spine.* 1997;22(10):1115-1122.
98. Kilpikoski S, Airaksinen O, Kankaanpaa M et al. Interexaminer reliability of low back pain assessment using the McKenzie method. *Spine.* 2002;27(8):E207-E214.
99. Chung SS, Lee CS, Kim SH et al. Effect of low back posture on the morphology of the spinal canal. *Skeletal Radiol.* 2000;29(4):217-223.
100. Zamani AA, Moriarty T, Hsu L et al. Functional MRI of the lumbar spine in erect position in a superconducting open-configuration MR system: preliminary results. *J Magn Reson Imaging.* 1998;8(6):1329-1333.
101. Lappalainen AK, Kaapa E, Lamminen A et al. The diagnostic value of contrast-enhanced magnetic resonance imaging in the detection of experimentally induced anular tears in sheep. *Spine.* 2002;27(24):2806-2810.
102. Ross JS, Modic MT, Masaryk TJ. Tears of the anulus fibrosus: assessment with Gd-DTPA-enhanced MR imaging. *AJR Am J Roentgenol.* 1990;154(1):159-162.
103. Snow RD, Williams JP, Weber ED et al. Enhancing transdural lumbar disk herniation. *Clin Imaging.* 1995;19(1):12-16.
104. Yoshida H, Fujiwara A, Tamai K et al. Diagnosis of symptomatic disc by magnetic resonance imaging: T2-weighted and gadolinium-DTPA-enhanced T1-weighted magnetic resonance imaging. *J Spinal Disord Tech.* 2002;15(3):193-198.
105. Thelander U, Fagerlund M, Friberg S et al. Straight leg raising test versus radiologic size, shape, and position of lumbar disc hernias. *Spine.* 1992;17(4):395-399.
106. Fagerlund MK, Thelander U, Friberg S. Size of lumbar disc hernias measured using computed tomography and related to sciatic symptoms. *Acta Radiol.* 1990;31(6):555-558.

APPENDIX A: THE INTERVERTEBRAL DISC – CONTEXTUAL INFORMATION

In order to explain the mechanisms by which the disc may cause LBP, and which MRI findings are of interest, a brief review of the intervertebral disc's anatomy and pathophysiology is useful.

ANATOMY OF THE INTERVERTEBRAL DISC

The disc is composed of a gelatinous nucleus contained laterally by the collagenous lamellae of the annulus and superiorly and inferiorly by the cartilaginous endplates. The disc's nerve supply is abundant in the outer third of the annulus and limited or nonexistent in the middle third.[1-3] Thus, disc pain is generally felt only after a pathological mechanism affects the innervated peripheral parts, highlighting the need to focus on degenerative findings affecting the intervertebral disc's periphery.[4]

PATHOANATOMY OF THE INTERVERTEBRAL DISC

When the intervertebral disc degenerates, a sequence of morphological changes is observed, which involve the dessication of the disc content and affect the disc's capacity to transfer loads.[5] As a result, circumferential tears occur in the annulus fibrosus followed by radial tears, leading to a loss of restraint of the nucleus pulposus.[6,7] This condition, combined with the reduced hydration of the nucleus, can result in disc-space narrowing. The migration of the nucleus through tears may result in disc contour abnormalities such as bulging or herniation.[8] If the annulus ruptures completely and the nuclear material extrudes, the disc contour abnormality is considered non-contained.[9] In response to a perturbation in the disc's load transfer capacity reactive bone formation frequently occurs at both the margins of the vertebral body and the facet joints leading to the observation of osteophytes and stenosis at imaging.[10-13] The Combined Task Forces of the North American Spine Society, American Society of Spine Radiology, and American Society of Neuroradiology recommended the definitions of disc pathology for use in anatomy and imaging studies which are listed in Appendix B.[9]

REFERENCES - APPENDIX A

1. Malinsky J. The ontogenetic development of nerve terminations in the intervertebral discs of man. *Acta Anatomica.* 1959;38:96-113.
2. Jackson HC, Winkelmann RK, Bickel WH. Nerve endings in the human lumbar spinal column and related structures. *J Bone Joint Surg Am.* 1966;48(7):1272-1281.
3. Freemont AJ, Peacock TE, Goupille P et al. Nerve ingrowth into diseased intervertebral disc in chronic back pain. *Lancet.* 1997;350(9072):178-181.
4. Kuslich SD, Ulstrom CL, Michael CJ. The tissue origin of low back pain and sciatica: a report of pain response to tissue stimulation during operations on the lumbar spine using local anesthesia. *Orthop Clin North Am.* 1991;22(2):181-187.
5. McKenzie R, May S. The Intervertebral Disc. In: Writer's Ink, editor. *The Lumbar Spine Mechanical Diagnosis & Therapy*. Waikanae, New Zealand: Spinal Publications New Zealand Ltd, 2003: 67-86.
6. Yu S, Haughton VM, Sether LA et al. Criteria for classifying normal and degenerated lumbar intervertebral disks. *Radiology.* 1989;170(2):523-526.
7. Kumpan W, Salomonowitz E, Seidl G et al. The intravertebral vacuum phenomenon. *Skeletal Radiol.* 1986;15(6):444-447.
8. Beattie P. The relationship between symptoms and abnormal magnetic resonance images of lumbar intervertebral disks. *Phys Ther.* 1996;76(6):601-608.
9. Fardon DF, Milette PC. Nomenclature and classification of lumbar disc pathology. Recommendations of the Combined task Forces of the North American Spine Society, American Society of Spine Radiology, and American Society of Neuroradiology. *Spine.* 2001;26(5):E93-E113.
10. Yong-Hing K, Kirkaldy-Willis WH. The pathophysiology of degenerative disease of the lumbar spine. *Orthop Clin North Am.* 1983;14(3):491-504.
11. McCall I, O'Brien J. Plain films of the lumbar spine. In: Wiesel SW, Weinstein JN, Herkowitz H, Dvorak J, Bell G, editors. *The lumbar spine*. Philadelphia, PA: Saunders, 1996: 333-352.
12. Kirkaldy-Willis WH, Wedge JH, Yong-Hing K et al. Pathology and pathogenesis of lumbar spondylosis and stenosis. *Spine.* 1978;3(4):319-328.
13. Andersson GB. The Epidemiology of Spinal Disorders. In: Frymoyer JW, editor. *The Adult Spine: Principles and Practice*. Philadelphia, PA: Lippincott-Raven, 1997: 93-141.

APPENDIX B: DEFINITIONS OF DISC-RELATED IMAGING FINDINGS FROM THE NOMENCLATURE AND CLASSIFICATION OF LUMBAR DISC PATHOLOGY

Recommendations of the Combined task Forces of the North American Spine Society, American Society of Spine Radiology, and American Society of Neuroradiology

DISC CONTOUR ABNORMALITIES :

"bulging disc, bulge (n), bulge (v) : 1. A disc in which the contour of the outer anulus extends, or appears to extend, in the horizontal (axial) plane beyond the edges of the disc space, usually over greater than 50% (180 degrees) of the circumference of the disc [...] Bulging may or may not represent pathologic finding, physiologic variant, or normalcy".[1]

"herniated disc, herniation (n), herniate (v): 1. Localized displacement of disc material beyond the normal margins of the intervertebral disc space. [...] Localized means, by way of convention, less than 50% (180 degrees) of the circumference of the disc. [...] The normal margins of the intervertebral disc space are defined, craniad and caudad, by the vertebral body endplates and peripherally by the edges of the vertebral body ring apophyses, exclusive of osteophytic formations. [...] Herniated discs in the horizontal (axial) plane may be further subcategorized as protruded or extruded".[1]

"protruded disc, protrusion (n), protrude (v): 1. A herniated disc in which the greatest plane, in any direction, between the edges of the disc material beyond the disc space is less than the distance between the edges of the base, when measured in the same plane. [...] The test of protrusion is that there must be a localized (less than 50% or 180° of the circumference of the disc) displacement of disc tissue so that the distance between the corresponding edges of the displaced portion must not be greater than the distance between the edges of the base".[1]

"extruded disc, extrusion (n), extrude (v): A herniated disc in which, in at least one plane, any one distance between the edges of the disc material beyond the disc space is greater

than the distance between the edges of the base in the same plane; or when no continuity exists between the disc material beyond the disc space and that within the disc space. [...] Displacement beyond the outer anulus of disc material with any distance between its edges greater than the distance between the edges of the base distinguishes extrusion from protrusion. [...] Extruded discs in which all continuity with the disc of origin is lost may be further characterized as *sequestrated*. Disc material displaced away from the site of extrusion may be characterized as *migrated*."[1]

OTHER DISC-RELATED IMAGING FINDINGS

"*radial fissure or tear:* Disruption of anular fibers extending from the nucleus outward toward the periphery of the anulus, usually in the craniad-caudad (vertical) plane, though, at times, with occasional horizontal (transverse) components. Note: Occasionally a radial fissure extends in the transverse plane to include avulsion of the outer layers of anulus from the apophyseal ring."[1]

"*concentric tear:* Tear or fissure of the anulus characterized by separation, or break, of anular fibers, in a plane roughly parallel to the curve of the periphery of the disc, creating fluid-filled spaces between adjacent anular lamellae."[1]

"*osteophytes*: Focal hypertrophy of bone surface and/or ossification of soft tissue attachments to the bone."[1]

"*marginal osteophytes:* Osteophyte that protrudes from and beyond the outer perimeter of the vertebral endplate apophysis."[1]

"*disc space height:* The distance between the planes of the endplates of the vertebrae craniad and caudad to the disc."[1]

"*desiccated disc, disc with reduced signal intensity* 1. Disc with reduced water content, usually primarily of nuclear tissues. 2. Imaging manifestations of reduced water content of the disc; or apparent reduced water content, as from alterations in the concentration of hydrophilic glycosaminoglycans."[1]

REFERENCES - APPENDIX B

1. Fardon DF, Milette PC. Nomenclature and classification of lumbar disc pathology. Recommendations of the Combined task Forces of the North American Spine Society, American Society of Spine Radiology, and American Society of Neuroradiology. *Spine.* 2001;26(5):E93-E113.

APPENDIX C: AN ILLUSTRATED DESCRIPTION OF THE MANUAL SEGMENTATION PROCEDURE

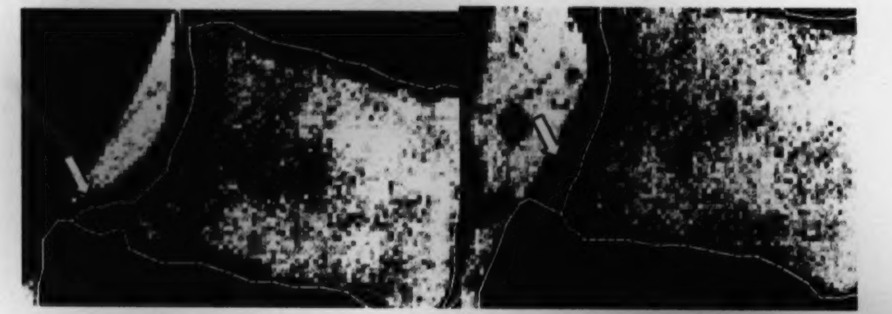
The following is a step by step description of the procedure used for manually segmenting the contour of the lumbar discs and vertebrae in order to derive the measures of disc contour abnormality studied in this dissertation. These segmentation guidelines were developed after examining outliers from a pilot reliability study.

For the sagittal slice of interest, the proton density image was first segmented as it displays the best overall contrast between the structures of interest. The brightness was adjusted to a value of 120 (different only if necessary to standardize). Contrast was adjusted by each user according to his/her preference to maximize the contrast between the structures of interest. Contrast was readjusted as necessary when segmenting different sections of the image.

Segmentation of the anterior vertebral wall and disc:

First, the anterior vertebral wall of the highest vertebra shown on the image was segmented from the anterior longitudinal ligament by a vertical line. At the disc level, the evaluator crossed the thickness of the anterior longitudinal ligament and continued by segmenting it from the soft tissue located in front of the spine. This was done by zooming the image by a factor of 5 and centering the image so that the anterior contour of the highest vertebra displayed on the image was shown. Then the evaluator began manual segmentation. The first point was placed as high on the image as possible at the interface between the bone and the anterior longitudinal ligament, then the line was continued by adding points along this interface as often as needed to follow the curvilinear contour of the bone.

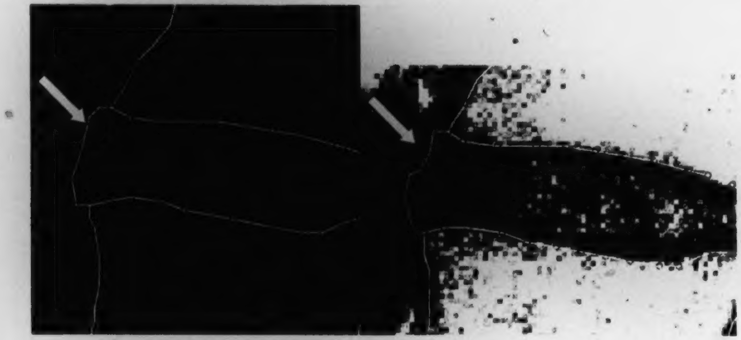
- If an osteophyte had started to form near the corners of the vertebra, its area was included within the vertebral contour only if the signal intensity at the junction between the osteophyte and the vertebra was similar to that of the osteophytes and vertebra. If a dark line of low signal intensity completely separated the osteophyte from the vertebra, the osteophyte area was not included within the vertebra area.



The last point on the vertebra contour was placed at the lower limit of the anterior wall. The next point was placed outside of the line of dark signal intensity that corresponded to the anterior longitudinal ligament in a direction parallel to a line that corresponded to the prolongement of the lower endplate of the vertebra. This was done because the contrast did not allow distinguishing the annulus from the longitudinal ligament.

Evaluators then continued adding points along the interface between the abdominal soft tissue and the anterior longitudinal ligament as often as needed to follow the curvilinear contour of the disc. Upon reaching the lower limit of the disc contour, a final point was placed at a level that corresponded to where a prolongment of the upper endplate would cross that interface. If disc material extended superiorly or inferiorly beyond the disc space, the evaluator continued adding points to include the disc material only if the signal intensity was homogeneous with that of the disc material within the center of the disc. The next point was placed at the highest limit of the anterior vertebral wall of the vertebra below the disc. The evaluator continued adding points along this interface as often as needed to follow the curvilinear contour of the bone. This was repeated for each disc and vertebra until reaching the lower limit of S1.

- Often times, at the disc level, it was difficult to distinguish the annulus/ligament from the large blood vessels at the front of the spine. When in doubt, the evaluators placed at least 3 points along the line of the anterior disc boundary that could be adjusted later when viewing the outline on other sequences providing better contrast for segmenting the ligament from blood vessels.

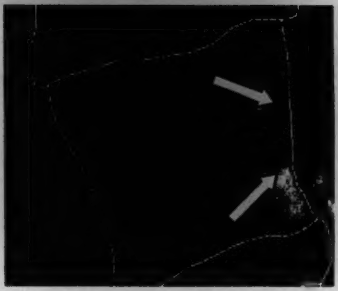


Segmentation of the posterior vertebral wall and disc:

The evaluators manually segmented the posterior vertebral wall from the posterior longitudinal ligament/submembranous material by creating a vertical line starting at the top of the proton density image continuing at the disc level crossing the thickness of the posterior longitudinal ligament and segmenting it from the soft tissue or CSF fluid located at the back of the spine.

See above for details on how disc material that had migrated was segmented. Because of the presence of fat in the sub-membranous space, it was sometimes difficult to distinguish migrated nuclear material from the fat in this area. Corrections were made on the T1-weighted image where the fat appeared bright and the nuclear material was relatively darker.

- The presence of the vertebral artery entering the vertebral body at mid-height created a challenge in manually segmenting the posterior vertebral wall. Thus, the evaluator segmented the upper third of the posterior wall and connected directly to the lower third without following the interface between the bone and the artery at the level of the middle third of the vertebra.



Segmentation of upper and lower disc/vertebral endplate interfaces

The evaluator segmented the upper and lower endplates of each disc by creating horizontal lines along the endplates. The last horizontal segmentation line was placed by following the endplate below S1.The evaluators began each line well outside of the disc space so that when crossing the vertical line segmenting the anterior vertebral wall and disc this horizontal line was aligned in the general plane of the endplate to be segmented. Each line was completed by avoiding to cross the vertical segmentation lines described above more than once. This was important in the presence of disc material that migrated upwards or downwards.

Adjustment of segmentation using T2 -and T1-weighted images

The regions of interest, corresponding to the areas enclosed within the vertical and horizontal segmentation lines, were copied from the PD image onto the corresponding T2-weighted image to adjust the segmented outline. The brightness and contrast of the T2-weighted image were adjusted to maximize the contrast between the structures of interest. On the T2-weighted image, the contrast between the CSF and the other tissues was excellent and the posterior vertical segmentation line described earlier was adjusted at the disc level taking advantage of the improved constrast on this sequence. The nucleus was also best seen on T2-weighted images. Thus, the segmentation of migrated nucleus material was corrected on T2-weighted images. Note: T2-weighted and PD images were acquired simultaneously and the position of the two images corresponded exactly.

The regions of interest were then copied onto the corresponding T1-weighted images to finish adjusting the segmented outline. Because the T1-weighted image was not acquired at the same time as the PD and T2 images, the position of the regions of interest sometimes needed to be adjusted by translation vertically or horizontally. On the T1-weighted image, the contrast between the bone or fat and the other tissues was excellent. The bone, the osteophytes, the Schmorl's defects contours were adjusted on this sequence taking advantage of the improved contrast. The contour of the migrated disc material could be distinguished from the submembranous tissue on this sequence and was adjusted on this sequence, if necessary. The final revision of the outline was copied back onto all sequences used in the segmentation and the position was adjusted, if necessary.

APPENDIX D: VARIANCE COMPONENTS FOR CHAPTER 2

Table i) Percentage of the total error variance in quantitative measures of anterior disc contour abnormality due to differences in repeated measures by the same evaluator (dataset is 3 segmentations by one evaluator at both resolutions)

			Variance components								% error
Disc Level	Distance vs. Area	Corners	Persons	Occasion	Resolution	PO	PR	OR	POR	Total Variance	due to Occasion
L1L2	Area	Real	1616.26	31.39	0.00	0.00	24.68	34.59	1429.21	3136.14	0.00
	Area	Theo	2834.79	0.00	0.00	0.00	0.00	0.00	1394.42	4229.20	0.00
	Distance	Real	2.58	0.07	0.00	0.37	0.00	0.00	1.76	4.78	7.80
	Distance	Theo	5.11	0.00	0.00	0.55	0.00	0.16	2.15	7.96	6.91
L2L3	Area	Real	2959.80	63.88	0.00	514.29	853.35	0.00	2563.71	6955.03	7.39
	Area	Theo	5523.22	0.00	0.00	543.08	735.01	15.31	2190.24	9006.87	6.03
	Distance	Real	2.58	0.01	0.00	0.35	0.38	0.00	1.93	5.25	6.58
	Distance	Theo	4.63	0.00	0.00	0.33	0.36	0.25	1.72	7.30	4.46
L3L4	Area	Real	1243.27	0.00	0.00	292.04	0.00	142.60	686.00	2363.90	12.35
	Area	Theo	3520.51	0.00	0.00	139.45	0.00	265.87	890.21	4816.03	2.90
	Distance	Real	1.59	0.00	0.00	0.33	0.00	0.21	1.17	3.30	10.13
	Distance	Theo	3.50	0.00	0.00	0.12	0.00	0.34	1.45	5.41	2.21
L4L5	Area	Real	2511.02	129.64	0.00	203.28	0.00	0.00	1603.49	4447.43	4.57
	Area	Theo	4484.42	0.00	0.00	178.36	72.71	27.84	1881.57	6644.89	2.68
	Distance	Real	2.83	0.13	0.00	0.20	0.00	0.00	1.88	5.04	4.04
	Distance	Theo	4.91	0.00	0.00	0.35	0.08	0.10	1.95	7.39	4.80
L5S1	Area	Real	2766.35	7.26	0.00	168.97	0.00	0.00	813.30	3755.87	4.50
	Area	Theo	6493.19	7.85	0.00	44.24	0.00	0.00	1071.64	7616.92	0.58
	Distance	Real	2.95	0.00	0.00	0.07	0.00	0.00	0.91	3.93	1.65
	Distance	Theo	4.30	0.01	0.00	0.00	0.07	0.00	0.95	5.33	0.00
									Overall min		0.00
									Overall max		12.35
									Overall median		3.01
									Overall mean		3.42

Table ii) Percentage of the total error variance in quantitative measures of posterior disc contour abnormality due to differences in repeated measures by the same evaluator (dataset is 3 segmentations by one evaluator at both resolutions)

			Variance components								% error
Disc Level	Distance vs. Area	Corners	Persons	Occasion	Resolution	PO	PR	OR	POR	Total Variance	Due to Occasion
L1L2	Area	Real	699.83	56.14	0.00	8.93	13.39	7.73	91.35	877.37	1.02
	Area	Theo	758.41	40.63	0.04	4.78	15.20	6.18	97.92	923.17	0.52
	Distance	Real	2.05	0.24	0.00	0.10	0.00	0.00	0.41	2.80	3.60
	Distance	Theo	2.52	0.17	0.03	0.26	0.01	0.00	0.29	3.29	8.05
L2L3	Area	Real	781.57	115.00	0.43	21.43	9.17	5.82	139.77	1073.19	2.00
	Area	Theo	748.61	113.51	0.00	0.98	0.00	6.83	176.51	1046.43	0.09
	Distance	Real	2.43	0.32	0.00	0.00	0.00	0.02	0.73	3.49	0.00
	Distance	Theo	3.32	0.30	0.00	0.06	0.00	0.03	0.64	4.34	1.36
L3L4	Area	Real	2115.05	134.59	0.00	88.29	3.66	53.96	233.23	2628.77	3.36
	Area	Theo	2012.23	157.48	0.00	130.41	33.91	26.52	173.01	2533.58	5.15
	Distance	Real	5.94	0.24	0.00	0.11	0.00	0.26	0.57	7.12	1.59
	Distance	Theo	6.07	0.23	0.00	0.12	0.04	0.19	0.43	7.08	1.67
L4L5	Area	Real	2427.66	194.87	5.73	109.19	63.58	54.08	280.16	3135.26	3.48
	Area	Theo	2328.05	224.97	24.74	95.57	51.45	47.30	293.85	3065.93	3.12
	Distance	Real	5.63	0.40	0.00	0.00	0.10	0.12	0.60	6.86	0.00
	Distance	Theo	5.15	0.49	0.03	0.09	0.00	0.21	0.58	6.54	1.32
L5S1	Area	Real	1834.37	17.89	0.00	2.08	0.00	18.14	374.66	2247.15	0.09
	Area	Theo	1942.96	23.31	0.00	1.25	0.00	0.25	382.49	2350.26	0.05
	Distance	Real	5.10	0.07	0.00	0.34	0.18	0.05	1.30	7.06	4.88
	Distance	Theo	4.76	0.11	0.00	0.40	0.17	0.02	1.35	6.80	5.93
									Overall min		0.00
									Overall max		12.35
									Overall median		3.01
									Overall mean		3.42

Table iii) Percentage of the total error variance in quantitative measures of anterior disc contour abnormality due to inter-evaluator differences and differences between low and high resolution images (dataset is 1 segmentation by each evaluator at both resolutions)

			Variance components								% error	
Disc Level	Dist. Vs Area	Reference corners	Person	Evaluator	Resolution	PE	PR	ER	PER	TOTAL	Due to Digitizers	Due to Resolution
L1L2	Area	Real	1457.04	0.00	0.00	145.01	5.91	0.00	712.43	2320.38	6.2	0.3
	Area	Theo	2737.79	119.41	56.08	183.19	112.98	0.00	612.07	3821.52	4.8	3.0
	Distance	Real	2.29	0.00	0.14	0.34	0.00	0.00	1.82	4.59	7.4	0.0
	Distance	Theo	4.93	0.25	0.00	1.00	0.16	0.00	1.56	7.91	12.7	2.1
L2L3	Area	Real	2017.72	6.93	0.00	25.32	246.69	8.24	2750.35	5055.26	0.5	4.9
	Area	Theo	4431.53	103.35	179.17	496.09	331.66	0.00	2120.35	7662.15	6.5	4.3
	Distance	Real	1.82	0.17	0.31	0.24	0.53	0.00	2.41	5.48	4.3	9.7
	Distance	Theo	3.74	0.22	0.00	0.26	0.60	0.06	1.49	6.36	4.0	9.4
L3L4	Area	Real	1771.49	0.00	161.40	509.59	288.88	0.00	530.66	3262.01	15.6	8.9
	Area	Theo	3372.33	138.24	0.00	336.02	186.38	9.11	455.57	4497.65	7.5	4.1
	Distance	Real	2.01	0.05	0.66	0.72	0.16	0.00	1.15	4.76	15.2	3.4
	Distance	Theo	4.00	0.09	0.09	0.21	0.06	0.02	0.95	5.42	3.9	1.2
L4L5	Area	Real	1980.99	0.00	113.80	1049.77	403.78	49.17	719.82	4317.32	24.3	9.4
	Area	Theo	4416.83	206.91	0.00	890.17	492.63	0.00	816.31	6822.84	13.0	7.2
	Distance	Real	2.55	0.03	0.40	0.83	0.50	0.00	1.38	5.71	14.6	8.7
	Distance	Theo	4.66	0.08	0.00	0.94	0.73	0.00	1.37	7.78	12.1	9.4
L5S1	Area	Real	2431.05	0.00	15.89	642.30	234.72	0.00	315.40	3639.36	17.6	6.4
	Area	Theo	5903.32	0.28	4.29	710.37	65.47	0.00	494.38	7178.11	9.9	0.9
	Distance	Real	2.58	0.00	0.05	0.94	0.22	0.00	0.42	4.21	22.3	5.2
	Distance	Theo	3.98	0.00	0.03	0.75	0.11	0.03	0.60	5.49	13.7	1.9
									Overall min		0.0	0.0
									Overall max		24.3	9.7
									Overall Median		5.0	2.0
									Overall Mean		7.1	3.2

Table iv) Percentage of the total error variance in quantitative measures of posterior disc contour abnormality due to inter-evaluator differences and differences between low and high resolution images (dataset is 1 segmentation by each evaluator at both resolutions)

			Variance components								% error	
Disc Level	Dist. Vs Area	Reference corners	Person	Evaluator	Resolution	PE	PR	ER	PER	TOTAL	Due to Digitizers	Due to Resolution
	Area	Real	693.74	57.70	0.00	0.00	0.00	1.45	123.32	876.22	0.0	0.0
	Area	Theo	716.49	87.91	0.00	0.00	4.82	0.00	158.63	967.86	0.0	0.5
	Distance	Real	1.98	0.35	0.18	0.12	0.08	0.05	0.44	3.20	3.7	2.4
L1L2	Distance	Theo	2.21	0.53	0.08	0.11	0.09	0.02	0.66	3.70	3.1	2.4
	Area	Real	804.76	78.11	0.00	0.00	0.00	18.97	191.85	1093.69	0.0	0.0
	Area	Theo	784.39	72.08	0.00	0.00	0.00	29.03	240.30	1125.80	0.0	0.0
	Distance	Real	2.53	0.44	0.21	0.14	0.07	0.15	0.53	4.07	3.5	1.6
L2L3	Distance	Theo	2.97	0.34	0.00	0.16	0.08	0.16	0.55	4.25	3.7	1.8
	Area	Real	2252.64	92.64	53.75	145.38	34.29	0.00	246.40	2825.09	5.1	1.2
	Area	Theo	2366.80	114.32	124.44	167.45	60.45	0.00	204.23	3037.70	5.5	2.0
	Distance	Real	5.96	0.60	0.05	0.27	0.02	0.00	0.58	7.49	3.6	0.3
L3L4	Distance	Theo	6.71	0.53	0.00	0.04	0.10	0.08	0.57	8.03	0.5	1.2
	Area	Real	2274.72	122.11	23.46	132.07	163.39	11.18	474.17	3201.09	4.1	5.1
	Area	Theo	2208.03	169.29	89.82	50.97	35.73	9.96	561.51	3125.30	1.6	1.1
	Distance	Real	5.28	0.35	0.00	0.18	0.22	0.00	1.13	7.16	2.6	3.0
L4L5	Distance	Theo	4.31	0.65	0.00	0.00	0.21	0.00	1.48	6.64	0.0	3.1
	Area	Real	1263.64	0.00	28.02	156.38	0.00	39.59	623.05	2110.68	7.4	0.0
	Area	Theo	1220.67	0.00	96.20	260.30	0.00	20.41	719.60	2317.18	11.2	0.0
	Distance	Real	3.92	0.00	0.00	0.12	0.00	0.00	2.83	6.86	1.7	0.0
L5S1	Distance	Theo	3.44	0.00	0.05	0.61	0.00	0.00	2.58	6.68	9.2	0.0
										Overall min	0.0	0.0
										Overall max	24.3	9.7
										Overall Median	5.0	2.0
										Overall Mean	7.1	3.2

Table v) Percentage of the total error variance in qualitative measures of anterior disc bulging and herniation due to intra- and inter-rater differences (dataset is 2 ratings by 2 raters on high resolution image)

		Variance components								% error	
Variable	Disc Level	Person	Rater	Occasion	PR	PO	RO	PRO	Total variance	due to Rater	due to Occasion
Bulging	L1L2	0.34	0.21	0.00	0.30	0.00	0.00	0.53	1.38	21.4	0.0
	L2L3	0.61	0.35	0.00	0.62	0.00	0.03	0.40	2.01	30.7	0.0
	L3L4	0.71	0.28	0.00	0.12	0.05	0.04	0.48	1.68	7.4	2.8
	L4L5	0.53	0.30	0.00	0.36	0.00	0.10	0.69	1.99	18.0	0.0
	L5S1	0.00	0.10	0.00	0.29	0.00	0.00	0.06	0.44	65.0	0.0
Herniation	L1L2	0.05	0.03	0.00	0.03	0.00	0.00	0.04	0.16	22.1	0.0
	L2L3	0.10	0.07	0.00	0.05	0.00	0.00	0.10	0.32	16.8	0.0
	L3L4	0.08	0.04	0.00	0.02	0.00	0.00	0.07	0.23	9.0	0.0
	L4L5	0.04	0.05	0.00	0.04	0.00	0.01	0.07	0.20	19.1	0.0
	L5S1	0.00	0.02	0.00	0.04	0.00	0.00	0.02	0.08	50.0	0.0
								Overall Min		0.0	0.0
								Overall Max		65.0	9.5
								Overall Median		17.5	0.0
								Overall Mean		19.6	1.5

Table vi) Percentage of the total error variance in qualitative measures of posterior disc bulging and herniation due to intra- and inter-rater differences (dataset is 2 ratings by 2 raters on high resolution image)

		Variance components								% error	
Variable	Disc Level	Person	Rater	Occasion	PR	PO	RO	PRO	Total variance	due to Rater	due to Occasion
Bulging	L1L2	0.35	0.00	0.00	0.02	0.00	0.05	0.48	0.91	2.3	0.5
	L2L3	0.70	0.00	0.00	0.35	0.06	0.02	0.38	1.51	23.1	4.0
	L3L4	1.11	0.09	0.01	0.58	0.16	0.00	0.43	2.38	24.3	6.9
	L4L5	1.60	0.00	0.00	0.00	0.00	0.00	0.95	2.54	0.0	0.0
	L5S1	1.26	0.00	0.08	0.28	0.00	0.00	0.95	2.57	10.8	0.0
Herniation	L1L2	0.03	0.00	0.00	0.01	0.00	0.01	0.03	0.09	11.2	2.7
	L2L3	0.06	0.00	0.00	0.03	0.00	0.00	0.03	0.12	22.0	4.1
	L3L4	0.10	0.00	0.00	0.03	0.02	0.00	0.04	0.18	17.1	9.5
	L4L5	0.13	0.00	0.00	0.03	0.00	0.01	0.07	0.24	10.7	0.0
	L5S1	0.11	0.00	0.01	0.02	0.00	0.00	0.05	0.19	11.1	0.0
								Overall Min		0.0	0.0
								Overall Max		65.0	9.5
								Overall Median		17.5	0.0
								Overall Mean		19.6	1.5

APPENDIX E: VARIANCE COMPONENTS FOR CHAPTER 3.

Table i) Percentage of the total error variance in quantitative measures of anterior disc contour abnormality due to inter-scanner differences and differences between repeated measures by the same evaluator (dataset is 2 segmentations by one evaluator on images from 2 different scanners)

			Variance components								% Error	
Disc Level	Distance vs Area	Corners	Person	Scanner	Occasion	PS	PO	SO	PSO	Total variance	Due to Scanner	Due to Occasion
L1L2	Area	Real	3131.60	126.94	14.96	624.11	0.00	0.00	1036.75	4934.37	12.6	0.0
	Area	Theo	8075.83	0.00	41.52	394.80	0.00	0.00	872.58	9384.73	4.2	0.0
	Distance	Real	6.15	0.33	0.04	1.34	0.28	0.00	1.54	9.68	13.9	2.9
	Distance	Theo	16.88	0.02	0.09	0.87	0.17	0.00	1.14	19.17	4.6	0.9
L2L3	Area	Real	4225.17	0.00	13.35	711.95	0.00	0.00	1256.03	6206.50	11.5	0.0
	Area	Theo	6620.79	0.00	15.09	783.64	0.00	0.00	1805.07	9224.59	8.5	0.0
	Distance	Real	4.43	0.00	0.00	0.37	0.00	0.00	1.44	6.23	5.9	0.0
	Distance	Theo	6.81	0.00	0.00	0.47	0.00	0.15	1.77	9.20	5.1	0.0
L3L4	Area	Real	2016.54	13.70	0.00	412.88	0.00	0.00	772.17	3215.28	12.8	0.0
	Area	Theo	4416.30	0.00	0.00	224.09	0.00	0.00	1030.12	5670.50	4.0	0.0
	Distance	Real	1.74	0.09	0.00	0.72	0.15	0.00	0.68	3.38	21.4	4.4
	Distance	Theo	3.94	0.02	0.00	0.44	0.00	0.00	0.94	5.35	8.3	0.0
L4L5	Area	Real	3990.65	88.51	0.00	564.23	0.00	0.00	744.65	5388.04	10.5	0.0
	Area	Theo	7851.46	0.00	0.00	969.10	0.00	0.00	1089.34	9909.90	9.8	0.0
	Distance	Real	4.81	0.15	0.00	0.31	0.00	0.00	0.97	6.25	4.9	0.0
	Distance	Theo	6.58	0.03	0.00	0.64	0.14	0.00	1.16	8.56	7.5	1.7
L5S1	Area	Real	8002.58	473.70	0.00	628.64	0.00	23.49	644.27	9772.69	6.4	0.0
	Area	Theo	14768.13	1112.30	7.98	1396.98	0.00	0.00	1013.49	18298.87	7.6	0.0
	Distance	Real	5.22	0.51	0.00	0.68	0.00	0.05	1.07	7.54	9.1	0.0
	Distance	Theo	9.02	0.75	0.00	0.57	0.00	0.00	1.06	11.40	5.0	0.0
									Overall min		4.0	0.0
									Overall max		21.4	4.4
									Overall median		7.9	0.0
									Overall mean		8.7	0.5

Table ii) Percentage of the total error variance in quantitative measures of posterior disc contour abnormality due to inter-scanner differences and differences between repeated measures by the same evaluator (dataset is 2 segmentations by one evaluator on images from 2 different scanners)

			Variance components								% Error	
Disc Level	Distance vs Area	Corners	Person	Scanner	Occasion	PS	PO	SO	PSO	Total Variance	Due to Scanner	Due to Occasion
L1L2	Area	Real	458.43	0.00	0.00	96.87	0.00	0.00	91.91	647.21	15.0	0.0
	Area	Theo	670.47	0.00	0.00	124.06	0.00	0.61	79.98	875.12	14.2	0.0
	Distance	Real	1.99	0.02	0.00	0.35	0.00	0.00	0.57	2.92	12.1	0.0
	Distance	Theo	2.81	0.00	0.00	0.64	0.22	0.00	0.27	3.93	16.1	5.6
L2L3	Area	Real	1097.23	9.04	3.12	72.32	30.62	3.87	88.54	1304.74	5.5	2.3
	Area	Theo	1465.43	15.23	0.00	60.47	20.46	0.00	94.24	1655.83	3.7	1.2
	Distance	Real	3.56	0.01	0.01	0.31	0.26	0.00	0.31	4.47	7.0	5.8
	Distance	Theo	5.03	0.01	0.01	0.28	0.19	0.00	0.32	5.84	4.8	3.3
L3L4	Area	Real	1263.36	0.00	0.00	130.51	8.87	2.07	96.25	1501.07	8.7	0.6
	Area	Theo	1715.82	0.00	6.96	132.47	20.48	0.00	119.03	1994.75	6.6	1.0
	Distance	Real	3.34	0.00	0.00	0.34	0.06	0.00	0.42	4.16	8.3	1.4
	Distance	Theo	4.31	0.00	0.00	0.43	0.16	0.04	0.37	5.30	8.1	2.9
L4L5	Area	Real	1636.88	0.00	0.00	158.46	52.72	0.00	182.83	2030.89	7.8	2.6
	Area	Theo	1806.44	0.00	1.76	234.61	33.55	0.00	168.80	2245.16	10.4	1.5
	Distance	Real	4.90	0.00	0.00	0.48	0.20	0.00	0.40	5.98	8.0	3.4
	Distance	Theo	5.88	0.00	0.02	0.42	0.06	0.00	0.45	6.84	6.2	0.9
L5S1	Area	Real	2816.79	4.92	0.00	131.29	71.03	7.71	349.93	3381.66	3.9	2.1
	Area	Theo	3994.63	9.57	0.00	123.06	29.46	0.00	366.53	4523.25	2.7	0.7
	Distance	Real	6.53	0.00	0.00	0.48	0.10	0.00	0.77	7.88	6.1	1.3
	Distance	Theo	7.90	0.00	0.00	0.18	0.00	0.00	0.92	9.00	2.1	0.0
									Overall min		2.1	0.0
									Overall max		16.1	5.8
									Overall median		7.4	1.3
									Overall mean		7.9	1.8

APPENDIX F: RATIONALE FOR THE HYPOTHESES OF CONVERGENT AND DIVERGENT VALIDITY

Table i) Rationale for the hypotheses of convergent and divergent validity presented in Chapter 5.

	Hypothesis	Rationale
H1	-Quantitative measures of posterior and anterior disc contour abnormality will have moderate correlations with qualitative measures of bulging, -and low correlation with ratings of endplate irregularities.	-Bulging and disc contour abnormality are measures of a similar construct: the amount of disc material located beyond the disc boundaries. Only moderate correlations are expected between quantitative measures of contour abnormality and qualitative measures of bulging because qualitative measures have sub-optimal reliability and restriction on the range. -From a previous study of the correlation amongst different qualitative ratings endplate ratings did not cluster with bulging ratings and thus are not expected to correlate with contour abnormalities.[1] Furthermore, some authors suggested that contour abnormalities occur when a genetic defect influences the disc annulus[2,3] while endplate irregularities are suspected to occur when different changes in the genetic profile leads to weakness of the endplate instead.[4]
H2	-Quantitative measures of disc contour abnormality, using theoretical corners (as if no remodeling of the vertebrae due to degeneration had occurred), will correlate moderately with osteophytes ratings, -and measures using real corners will have relatively lower correlations with qualitative osteophyte ratings.	- A higher correlation between osteophytes and measures using theoretical corners is expected because in the presence of osteophytes, a larger distance or area would be measured because the theoretical corners would be located relatively closer to the center of the disc than the real corners. Some authors have suggested an association between the occurrence of contour abnormalities (bulging or herniation) and osteophytes.[5-8] Assuming a common underlying pathological process is at play, measuring contour abnormality using the real corners as reference would lead to underestimating the effect of that process because real corners are located where the osteophytes end and none of the disc area enclosed between the osteophytes would be. -With the development of osteophytes conjointly to contour abnormality, the measure of disc contour using real corners may not change while the disc contour measured using theoretical corners is expected to increase with the presence of osteophytes.
H3	-Correlation coefficients between quantitative measures of contour abnormality and qualitative bulging for measures obtained from para-sagittal images will be equal or larger than correlation coefficients for measures obtained from midsagittal images. -Correlations with qualitative measures of endplate irregularities and osteophytes will be similar when para-sagittal or midsagittal images are used.	- A higher or equal correlation is expected is because correlation coefficients tend to be stronger when the variance of each measure correlated is larger. Larger contour abnormalities and bulges are reportedly more frequent on para-sagittal images[9,10], thereby increasing the variances on para-sagittal images and preventing attenuation of the coefficients which might be observed on midsagittal images. Findings are all small on the latter images. - The prevalence of endplate irregularities and osteophytes is expected to remain similar on the para-sagittal images, thus the pattern of the correlation with measures of contour abnormality should remain the same.

Table i) Rationale for the hypotheses of convergent and divergent validity presented in Chapter 5 (continued)

H4	-Areas and distances of quantitative posterior contour abnormality will correlate strongly together -and correlations will be low with quantitative measures of anterior contour abnormality. -The corresponding hypothesis also applies anteriorly.	- All quantitative measures of disc contour abnormalities are intended to measure the same construct using a different method. Correlations between the measures of distance and area should all correlate strongly to demonstrate they address the same construct. -Measures of disc contour abnormality in the same location (e.g., posteriorly at L4L5) are expected to correlate more strongly together than with measures or contour abnormality in a different location (e.g., anteriorly at L4L5). The correlations between measures of disc contour from different locations at the same level might still be significant but should be lower than measures at the same location. This is because contour changes at a given disc might be due to a common cause and appear similarly at any point of the disc circumference.
H5	-All quantitative measurement methods for contour abnormality (areas and distances using both types of reference corners) will correlate strongly together. -However, similar measurement methods will correlate slightly more together (such as area with area, using real corners with using real corners) than with different methods.	- Similarly as for H4, all quantitative measures of disc contour abnormalities are intended to measure the same construct using a different method. Correlations between the measures using the two types of corners (real and theoretical) should correlate strongly demonstrating that they address the same construct. - However, it is possible, in both H4 and H5, that the measurement method introduces a small amount of common construct-irrelevant variance. Thus, measures of disc contour with similar methods may correlate slightly more strongly (distance vs distance…) than with measures using a different method (real vs theoretical corners…). If differences are small between the strength of the correlations for measures using similar and different methods, it will demonstrate that all measures address the same construct and that little construct-irrelevant variance is represented by the measurements.
H6	-In subjects without moderate or severe qualitative ratings of bulging and without herniation, area and distance measures of disc contour abnormality will all correlate strongly together. -However, in subjects identified by qualitative raters as having a herniation, which is typically asymmetric in the sagittal plane, areas will correlate more strongly together than with distances.	- Subjects without herniations and with smaller qualitative grades of bulging are expected to have symmetrical disc contour abnormalities in the sagittal plane with respect to a line bisecting the disc at its midline.[9] Thus, measures of area and distance should increase proportionally with increasing degrees of disc abnormality. This proportional increase is why strong correlations between both types of measures are expected. - In contrast, subjects with herniation and larger grades of bulging are expected to have contour abnormalities that are asymmetrical in the sagittal plane with respect to a line bisecting the disc at its midline.[9] Disc material reportedly migrates vertically in one direction more than the other when herniation or large bulges are present.[9] The migrated disc material would be detected by an area measures of contour abnormality but not by a distance measure obtained at mid-disc height. This discrepancy would result in non-proportional increases in both types of the measures with the presence of migrated disc material and would reduce the correlations expected between both types of measures.

REFERENCES - APPENDIX F

1. Videman T, Battie MC, Gill K et al. Magnetic resonance imaging findings and their relationships in the thoracic and lumbar spine. Insights into the etiopathogenesis of spinal degeneration. *Spine* 1995;20:928-35.
2. Kawaguchi Y, Kanamori M, Ishihara H et al. The association of lumbar disc disease with vitamin-D receptor gene polymorphism. *J.Bone Joint Surg.Am.* 2002;84-A:2022-8.
3. Videman T, Leppavuori J, Kaprio J et al. Intragenic polymorphisms of the vitamin D receptor gene associated with intervertebral disc degeneration. *Spine* 1998;23:2477-85.
4. Karppinen J, Paakko E, Paassilta P et al. Radiologic phenotypes in lumbar MR imaging for a gene defect in the COL9A3 gene of type IX collagen. *Radiology* 2003;227:143-8.
5. Milette PC. Classification, diagnostic imaging, and imaging characterization of a lumbar herniated disk. *Radiol.Clin.North Am.* 2000;38:1267-92.
6. McCall I, O'Brien J. Plain films of the lumbar spine. In: Wiesel SW, Weinstein JN, Herkowitz H, Dvorak J, Bell G, eds. *The lumbar spine*. 2nd ed. Philadelphia, PA: Saunders, 1996:333-52.
7. Andersson GB. The Epidemiology of Spinal Disorders. In: Frymoyer JW, ed. *The Adult Spine: Principles and Practice*. 2nd ed. Philadelphia, PA: Lippincott-Raven, 1997:93-141.
8. Vernon-Roberts B and Pirie CJ. Degenerative changes in the intervertebral discs of the lumbar spine and their sequelae. *Rheumatol.Rehabil.* 1977;16:13-21.
9. Ebeling U and Reulen HJ. Are there typical localisations of lumbar disc herniations? A prospective study. *Acta Neurochir.(Wien.)* 1992;117:143-8.
10. Wittenberg RH, Lutke A, Longwitz D et al. The correlation between magnetic resonance imaging and the operative and clinical findings after lumbar microdiscectomy. *Int.Orthop.* 1998;22:241-4.

APPENDIX G: FREQUENCY DISTRIBUTIONS OF THE QUALITATIVE SCORES ATTRIBUTED BY THE DIFFERENT RATERS USING THE DIFFERENT RATING DEFINITIONS FOR THE DIFFERENT SAMPLES.

Table i) Percent Frequency Distributions of the Midsagittal Qualitative Scores Attributed by the Different Raters Using the Different Rating Definitions for the Different Samples.

Orthopedist using rating from Table 5.2 Sample 1					Orthopedist, ratings from Table 5.2, Sample 1b					Orthopedist, ratings from Table 5.3, Sample 1b					1st Radiologist, ratings from Table 5.3, Sample 1b					2nd Radiologist, ratings from Table 5.4, Sample 1b				
(n=132)	% with Score				(n=88)	% with Score				(n=88)	% with Score				(n=88)	% with Score				(n=88)	% with Score			
	0	1	2	3		0	1	2	3		0	1	2	3		0	1	2	3		0	1	2	3
										Anterior bulging														
L1L2	77	17	5	0	L1L2	81	14	6	0	L1L2	76	16	7	1	L1L2	76	16	7	1	L1L2	91	7	1	1
L2L3	63	27	9	1	L2L3	65	26	9	0	L2L3	77	13	8	2	L2L3	77	13	8	2	L2L3	72	17	10	1
L3L4	55	36	8	1	L3L4	56	35	9	0	L3L4	90	3	3	3	L3L4	90	3	3	3	L3L4	55	34	11	0
L4L5	37	42	17	3	L4L5	35	44	19	1	L4L5	91	2	3	3	L4L5	91	2	3	3	L4L5	32	44	23	1
L5S1	52	36	13	0	L5S1	39	45	16	0	L5S1	88	3	9	0	L5S1	88	3	9	0	L5S1	40	35	22	3
Mean ant. bulging	**57**	**32**	**11**	**1**		**55**	**33**	**12**	**0**		**84**	**8**	**6**	**2**		**84**	**8**	**6**	**2**		**58**	**28**	**13**	**1**
										Posterior bulging														
L1L2	82	14	5	0	L1L2	85	10	5	0	L1L2	95	3	1	0	L1L2	95	3	1	0					
L2L3	67	24	8	1	L2L3	67	25	8	0	L2L3	92	8	0	0	L2L3	92	8	0	0					
L3L4	53	39	8	1	L3L4	57	35	8	0	L3L4	86	13	1	0	L3L4	86	13	1	0					
L4L5	28	46	23	3	L4L5	30	48	22	1	L4L5	77	17	6	0	L4L5	77	17	6	0					
L5S1	36	49	15	0	L5S1	34	49	17	0	L5S1	66	28	6	0	L5S1	66	28	6	0					
Mean post. bulging	**53**	**34**	**12**	**1**		**55**	**33**	**12**	**0**		**83**	**14**	**3**	**0**		**83**	**14**	**3**	**0**					
										Endplates irregularities														
L1L2	80	11	7	2	L1L2	82	10	6	2	L1L2	61	16	19	3	L1L2	61	16	19	3					
L2L3	80	11	7	2	L2L3	84	7	8	1	L2L3	69	11	16	3	L2L3	69	11	16	3					
L3L4	89	6	4	1	L3L4	86	8	5	1	L3L4	67	14	17	2	L3L4	67	14	17	2					
L4L5	91	4	5	0	L4L5	92	3	5	0	L4L5	70	18	7	5	L4L5	70	18	7	5					
L5S1	95	2	2	1	L5S1	97	2	1	0	L5S1	68	7	17	8	L5S1	68	7	17	8					
Mean end. Irreg.	**87**	**7**	**5**	**1**		**88**	**6**	**5**	**1**		**67**	**13**	**15**	**4**		**67**	**13**	**15**	**4**					
										Osteophytes														
L1L2	66	20	11	2	L1L2	69	17	13	1	L1L2	86	0	14	0	L1L2	86	0	14	0					
L2L3	55	26	17	2	L2L3	56	25	17	2	L2L3	85	0	14	1	L2L3	85	0	14	1					
L3L4	62	22	11	5	L3L4	65	20	11	3	L3L4	81	0	16	3	L3L4	81	0	16	3					
L4L5	58	22	14	7	L4L5	63	22	14	2	L4L5	82	0	16	2	L4L5	82	0	16	2					
L5S1	71	22	6	1	L5S1	69	25	6	0	L5S1	85	0	15	0	L5S1	85	0	15	0					
Mean osteophytes	**62**	**22**	**12**	**3**		**64**	**22**	**12**	**2**		**84**	**0**	**15**	**1**		**84**	**0**	**15**	**1**					
Mean all	**65**	**24**	**10**	**2**		**66**	**24**	**10**	**1**		**80**	**9**	**10**	**2**		**80**	**9**	**10**	**2**		**58**	**28**	**13**	**1**

Table ii) Frequency Distributions of the Qualitative Scores Attributed by the Orthopedist Using Definitions from Table 5.2 for the Midsagittal, Right and Left Images in Sample 2

	Midsagittal images % with score					Right para-sagittal images % with score					Left para-sagittal images % with score			
	0	1	2	3		0	1	2	3		0	1	2	3
						Anterior bulging								
L2L3	62	26	9	4	L2L3	55	38	4	2	L2L3	53	40	4	2
L3L4	49	32	13	4	L3L4	36	47	13	2	L3L4	36	47	13	2
L4L5	45	36	19	0	L4L5	28	49	23	0	L4L5	28	49	23	0
L5S1	68	26	6	0	L5S1	40	47	13	0	L5S1	40	47	13	0
Mean Ant. bulging	**56**	**30**	**12**	**2**		**40**	**45**	**13**	**1**		**39**	**46**	**13**	**1**
						Posterior bulging								
L2L3	66	26	6	2	L2L3	55	38	4	2	L2L3	55	38	4	2
L3L4	0	83	15	2	L3L4	36	47	13	2	L3L4	36	47	13	2
L4L5	28	57	15	0	L4L5	28	49	23	0	L4L5	28	49	23	0
L5S1	43	49	9	0	L5S1	40	47	13	0	L5S1	40	47	13	0
Mean Post. Bulging	**34**	**54**	**11**	**1**		**40**	**45**	**13**	**1**		**40**	**45**	**13**	**1**
	Endplate irregularities													
L2L3	83	13	4	0										
L3L4	87	9	4	0										
L4L5	94	6	0	0										
L5S1	100	0	0	0										
Mean End. Irreg.	**91**	**7**	**2**	**0**										
	Osteophytes													
L2L3	47	30	17	6										
L3L4	47	36	13	4										
L4L5	45	34	17	4										
L5S1	74	19	6	0										
Mean Osteophytes	**53**	**30**	**13**	**4**										
Mean all	**59**	**30**	**9**	**2**		**40**	**45**	**13**	**1**		**40**	**45**	**13**	**1**

APPENDIX H: MEANS OF THE DIFFERENT QUANTITATIVE MEASURES OF DISC CONTOUR ABNORMALITY FOR EACH SAMPLE

Table i) Mean Quantitative Measures of Disc Contour Abnormality for the Midsagittal Image of Subjects in Sample 1

N = 132	Anterior				Posterior			
	Real		Theoretical		Real		Theoretical	
	Area (mm^2)	Distance (mm)	Area (mm^2)	Distance (mm)	Area (mm^2)	Distance (mm)	Area (mm^2)	Distance (mm)
L1L2	69 ± 47	4 ± 2	144 ± 72	7 ± 3	49 ± 28	3 ± 2	65 ± 29	5 ± 2
L2L3	101 ± 68	5 ± 3	188 ± 86	8 ± 3	68 ± 45	4 ± 2	88 ± 45	6 ± 2
L3L4	109 ± 74	5 ± 2	196 ± 90	8 ± 3	79 ± 40	5 ± 2	100 ± 42	6 ± 2
L4L5	132 ± 83	5 ± 2	255 ± 116	9 ± 3	101 ± 51	6 ± 2	122 ± 50	8 ± 2
L5S1	132 ± 90	5 ± 3	327 ± 131	12 ± 3	104 ± 70	7 ± 3	114 ± 73	8 ± 3

Table ii) Mean Quantitative Measures of Disc Contour Abnormality for the Midsagittal Image of Subjects in Sample 1b

N = 88	Anterior				Posterior			
	Real		Theoretical		Real		Theoretical	
	Area (mm^2)	Distance (mm)	Area (mm^2)	Distance (mm)	Area (mm^2)	Distance (mm)	Area (mm^2)	Distance (mm)
L1L2	74 ± 45	4 ± 2	139 ± 61	7 ± 3	51 ± 23	4 ± 2	65 ± 25	5 ± 2
L2L3	114 ± 69	5 ± 2	187 ± 88	8 ± 3	72 ± 43	4 ± 2	90 ± 44	6 ± 2
L3L4	120 ± 70	5 ± 2	193 ± 84	8 ± 3	81 ± 39	5 ± 2	104 ± 43	6 ± 2
L4L5	138 ± 72	6 ± 2	249 ± 92	9 ± 3	104 ± 52	6 ± 3	125 ± 50	8 ± 2
L5S1	146 ± 83	6 ± 3	335 ± 126	12 ± 3	111 ± 75	7 ± 3	121 ± 79	8 ± 3

Table iii) Mean Quantitative Measures of Disc Contour Abnormality for the Midsagittal Image of Subjects in Sample 2

N = 47	Anterior				Posterior			
	Real		Theoretical		Real		Theoretical	
	Area (mm^2)	Distance (mm)	Area (mm^2)	Distance (mm)	Area (mm^2)	Distance (mm)	Area (mm^2)	Distance (mm)
L1L2	105 ± 75	5 ± 3	195 ± 91	9 ± 3	66 ± 27	4 ± 2	81 ± 29	5 ± 2
L2L3	119 ± 71	5 ± 2	214 ± 86	9 ± 3	84 ± 34	5 ± 2	101 ± 33	6 ± 2
L3L4	142 ± 76	6 ± 3	244 ± 95	9 ± 3	106 ± 38	6 ± 2	129 ± 42	8 ± 2
L4L5	159 ± 87	6 ± 3	278 ± 98	10 ± 3	137 ± 51	8 ± 2	158 ± 55	9 ± 2
L5S1	138 ± 86	5 ± 3	326 ± 108	12 ± 3	123 ± 65	8 ± 3	135 ± 73	8 ± 3

Table iv) Mean Quantitative Measures of Disc Contour Abnormality for the Right Para-sagittal Images of Subjects in Sample 2

N = 47	Anterior				Posterior			
	Real		Theoretical		Real		Theoretical	
	Area (mm^2)	Distance (mm)	Area (mm^2)	Distance (mm)	Area (mm^2)	Distance (mm)	Area (mm^2)	Distance (mm)
L1L2	171 ± 119	8 ± 4	298 ± 131	13 ± 5	94 ± 37	5 ± 2	111 ± 37	6 ± 2
L2L3	234 ± 161	8 ± 3	371 ± 156	14 ± 4	124 ± 43	6 ± 2	148 ± 45	8 ± 2
L3L4	236 ± 124	8 ± 3	376 ± 146	13 ± 4	136 ± 45	7 ± 2	164 ± 46	9 ± 2
L4L5	216 ± 111	7 ± 3	352 ± 128	12 ± 3	147 ± 49	8 ± 2	170 ± 46	10 ± 2
L5S1	161 ± 88	6 ± 3	365 ± 117	13 ± 3	145 ± 75	9 ± 3	160 ± 83	9 ± 3

Table v) Mean Quantitative Measures of Disc Contour Abnormality for the Left Para-sagittal Images of Subjects in Sample 2

N = 47	Anterior				Posterior			
	Real		Theoretical		Real		Theoretical	
	Area (mm^2)	Distance (mm)	Area (mm^2)	Distance (mm)	Area (mm^2)	Distance (mm)	Area (mm^2)	Distance (mm)
L1L2	189 ± 123	8 ± 4	296 ± 138	12 ± 5	86 ± 29	5 ± 2	101 ± 29	6 ± 1
L2L3	225 ± 129	8 ± 4	346 ± 144	13 ± 4	119 ± 45	6 ± 2	141 ± 45	8 ± 2
L3L4	264 ± 158	9 ± 4	389 ± 173	13 ± 5	134 ± 39	7 ± 2	160 ± 41	9 ± 2
L4L5	255 ± 136	8 ± 3	397 ± 170	13 ± 3	147 ± 48	8 ± 2	177 ± 52	10 ± 2
L5S1	183 ± 85	7 ± 3	379 ± 119	14 ± 3	153 ± 83	9 ± 3	168 ± 86	10 ± 3

Table vi) Mean Quantitative Measures of Disc Contour Abnormality for the Midsagittal Image of Subjects in Sample 3

N = 287	Anterior				Posterior			
	Real		Theoretical		Real		Theoretical	
	Area (mm^2)	Distance (mm)	Area (mm^2)	Distance (mm)	Area (mm^2)	Distance (mm)	Area (mm^2)	Distance (mm)
L1L2	100 ± 70	5 ± 3	158 ± 90	8 ± 4	57 ± 26	4 ± 2	72 ± 28	5 ± 2
L2L3	138 ± 84	7 ± 3	205 ± 100	9 ± 3	77 ± 38	5 ± 2	97 ± 40	6 ± 2
L3L4	150 ± 93	6 ± 3	220 ± 108	9 ± 3	92 ± 42	6 ± 2	114 ± 45	7 ± 2
L4L5	163 ± 86	6 ± 3	259 ± 100	10 ± 3	112 ± 49	7 ± 3	130 ± 50	8 ± 3
L5S1	188 ± 92	7 ± 3	376 ± 114	13 ± 3	109 ± 60	7 ± 3	119 ± 62	8 ± 3